

UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS POLÍTICAS Y SOCIOLOGÍA



TESIS DOCTORAL

**¿Para qué sirven las evaluaciones a los programas
sociales? La utilización y el efecto de las evaluaciones a los
programas de la Secretaría de Desarrollo Social en México,
2007-2014**

MEMORIA PARA OPTAR AL GRADO DE DOCTORA

PRESENTADA POR

Beatriz Carolina Corral Osuna

DIRECTORA

Laura Flamand Gómez

Madrid, 2018



UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS POLÍTICAS Y SOCIOLOGÍA
INSTITUTO UNIVERSITARIO DE INVESTIGACIÓN
ORTEGA Y GASSET

GOBIERNO Y ADMINISTRACIÓN PÚBLICA

Tesis Doctoral

**¿PARA QUÉ SIRVEN LAS EVALUACIONES A LOS PROGRAMAS SOCIALES?
LA UTILIZACIÓN Y EL EFECTO DE LAS EVALUACIONES A LOS PROGRAMAS
DE LA SECRETARÍA DE DESARROLLO SOCIAL EN MÉXICO, 2007-2014**

BEATRIZ CAROLINA CORRAL OSUNA

Director de la Tesis

**DRA. LAURA FLAMAND GÓMEZ
EL COLEGIO DE MÉXICO**

Madrid, 2017

AGRADECIMIENTOS

Dedico esta tesis a mi querida familia. Especialmente a mi madre y a mi hermana, Isabel y Miriam, que como siempre me animaron a seguir mis sueños y a concluir esta etapa. Por brindarme su amor incondicional y estar siempre cerca de mí sin importar la distancia, las quiero mucho.

A La Agencia Española de Cooperación Internacional para el Desarrollo (AECID) que generosamente me abrió las puertas de España y me brindó una experiencia de vida que trascendió lo académico. Por financiar mi estancia en Madrid y mi formación doctoral en el Instituto Universitario de Investigación Ortega y Gasset (IUIOG).

A mi Directora de Tesis, la Dra. Laura Flamand, quien creyó en mí e hizo posible este trabajo gracias a su paciencia y sus consejos. Por desempeñar su papel con toda la seriedad requerida sin dejar de lado su calidez humana.

A los miembros del tribunal quienes accedieron a estudiar mi tesis y la aprobaron.

A todos mis amigos y seres queridos que me acompañaron durante este largo proceso, en especial a aquellos que ya no pudieron compartir conmigo este día: Federico Sánchez Scott y Antonio Herrejón. Todo mi cariño y agradecimiento a donde quiera que estén.

Y por último, pero no menos importante, a Miguel y José Luis por leerme, escucharme e incluso consolarme en momentos difíciles.

ÍNDICES

ÍNDICE GENERAL

ÍNDICE DE CUADROS	VIII
ÍNDICE DE GRÁFICOS	X
ÍNDICE DE ILUSTRACIONES	XII
SIGLAS Y ACRÓNIMOS	XVI
DEFINICIONES BÁSICAS Y GLOSARIO DE TÉRMINOS	XX
RESUMEN	XXIV
ABSTRACT.....	XXX

INTRODUCCIÓN	2
JUSTIFICACIÓN PRÁCTICA	5
JUSTIFICACIÓN TEÓRICA	9
OBJETIVO GENERAL	18
PREGUNTAS DE INVESTIGACIÓN	18
HIPÓTESIS	19

CAPÍTULO 1. REVISIÓN DE LA LITERATURA SOBRE EL APROVECHAMIENTO DE LAS	
EVALUACIONES: ANÁLISIS DE LOS FACTORES DETERMINANTES Y LOS EFECTOS.....	22
1.1. EL DESARROLLO DE LOS DEBATES SOBRE EL APROVECHAMIENTO DE LA EVALUACIÓN	25
1.1.1. EL USO (1971 – 1985)	26
1.1.2. LA UTILIZACIÓN (1986 – 2000)	29
1.1.3. LA INFLUENCIA (2000 – PRESENTE).....	32
1.2. ¿QUÉ FACTORES DETERMINAN SI LAS EVALUACIONES SE UTILIZAN?	35
1.2.1. FACTORES ENDÓGENOS: LAS CARACTERÍSTICAS DE LA EVALUACIÓN	38
1.2.2. FACTORES INSTITUCIONALES: LA INDEPENDENCIA DEL PROCESO EVALUATIVO	44
1.3. ¿CUÁL ES EL EFECTO DE LA EVALUACIÓN EN LAS INTERVENCIONES PÚBLICAS?	48
1.4. EL APROVECHAMIENTO DE LAS EVALUACIONES EN MÉXICO	53
1.4.1. ANÁLISIS DE FACTORES QUE EXPLICAN LA UTILIZACIÓN	54
1.4.2. ANÁLISIS DEL EFECTO DE LAS RECOMENDACIONES	59
REFLEXIONES FINALES.....	62

CAPÍTULO 2. PANORAMA INICIAL SOBRE LAS EVALUACIONES REALIZADAS A LOS PROGRAMAS DE	
SEDESOL Y SU UTILIZACIÓN, 2007-2014.....	66

2.1. LA PRÁCTICA DE EVALUACIÓN EN SEDESOL.....	68
2.1.1. ORGANISMOS Y EXPERTOS EVALUADORES	73
2.1.2. EL COSTO	77
2.1.3. LA FUENTE DE FINANCIAMIENTO	80
2.2. LA UTILIZACIÓN DE LAS EVALUACIONES	82
2.2.1. DE RECOMENDACIONES A ASPECTOS SUSCEPTIBLES DE MEJORA.....	84
2.2.2. LA ATENCIÓN DE LOS ASPECTOS SUSCEPTIBLES DE MEJORA	90
REFLEXIONES FINALES.....	93

CAPÍTULO 3. LA IMPORTANCIA DE LAS CARACTERÍSTICAS DE LAS EVALUACIONES Y DE LA INDEPENDENCIA DEL PROCESO EVALUATIVO EN LA UTILIZACIÓN	98
3.1 ASPECTOS GENERALES SOBRE LA REALIZACIÓN Y LA UTILIZACIÓN DE LAS EVALUACIONES	100
3.1.1. ¿QUÉ CRITERIOS SE ATIENDEN EN LA REALIZACIÓN DE LAS EVALUACIONES?.....	101
3.1.2. ¿CÓMO SE PROMUEVE LA IMPLEMENTACIÓN DE LAS RECOMENDACIONES?	107
3.2. LA RELEVANCIA DE FACTORES ENDÓGENOS: ¿LOS ATRIBUTOS DE LA EVALUACIÓN AYUDAN A EXPLICAR SU UTILIZACIÓN?.....	110
3.2.1. ANÁLISIS DE DATOS.....	113
3.2.2. RESULTADOS: REGRESIÓN MULTIVARIADA DE MÍNIMOS CUADRADOS ORDINARIOS	117
3.3. LA RELEVANCIA DE FACTORES EXÓGENOS: ¿LA INDEPENDENCIA DEL PROCESO EVALUATIVO DETERMINA LA UTILIZACIÓN DE LAS RECOMENDACIONES?	120
3.3.1. ANÁLISIS DE DATOS.....	126
3.3.2. RESULTADOS: REGRESIÓN MULTIVARIADA DE MÍNIMOS CUADRADOS ORDINARIOS	130
REFLEXIONES FINALES.....	133

CAPÍTULO 4. LA UTILIZACIÓN INSTRUMENTAL DE LAS RECOMENDACIONES DERIVADAS DE LAS EVALUACIONES Y SU EFECTO EN EL MEJORAMIENTO DE LOS PROGRAMAS	138
4.1. EL EFECTO DIRECTO EN EL DISEÑO: CAMBIOS EN LAS REGLAS DE OPERACIÓN	141
4.1.1. EL EFECTO DE LAS EVALUACIONES EN EL DISEÑO DE LOS PROGRAMAS	150
4.2. EL EFECTO NO INMEDIATO EN EL DESEMPEÑO ¿LAS RECOMENDACIONES AFECTAN EL INDEP DE LOS PROGRAMAS?.....	153
4.2.1. EL ÍNDICE DE DESEMPEÑO DE LOS PROGRAMAS PÚBLICOS FEDERALES (INDEP)	154
4.2.2. CAMBIOS EN EL INDEP A PARTIR DE LA UTILIZACIÓN DE RECOMENDACIONES.....	159
REFLEXIONES FINALES.....	169

CAPÍTULO 5. LA UTILIZACIÓN Y EL EFECTO DE LAS EVALUACIONES: ESTUDIO DE CASO DE PDZP (2010 Y 2014) Y 3x1 (2008 Y 2014).....	174
5.1. ESTUDIOS DE CASO SELECCIONADOS	175
5.1.1. PROGRAMA PARA EL DESARROLLO DE ZONAS PRIORITARIAS (PDZP), 2009 – 2014.....	178
5.1.2. PROGRAMA 3x1 PARA MIGRANTES (3x1), 2007 - 2014.....	180
5.2. LA UTILIZACIÓN DE LAS EVALUACIONES	182
5.2.1. FORMALIZACIÓN E IMPLEMENTACIÓN DE LOS ASM.....	183
5.3. EL EFECTO DE LAS EVALUACIONES EN LOS PROGRAMAS.....	192
5.3.1. LOS CAMBIOS EN EL DISEÑO	194
5.3.2. LOS CAMBIOS EN EL DESEMPEÑO.....	199
REFLEXIONES FINALES.....	201
 CAPÍTULO 6. CONCLUSIONES	 206
6.1. SOBRE EL ALCANCE DE LA PRÁCTICA EVALUATIVA	207
6.2. LA UTILIZACIÓN Y EL EFECTO DE LAS EVALUACIONES: BENEFICIOS LIMITADOS EN LOS PROGRAMAS SOCIALES	210
6.3. SOBRE LOS RETOS IDENTIFICADOS EN LA PRÁCTICA EVALUATIVA ORIENTADA A PROGRAMAS SOCIALES.....	214
 BIBLIOGRAFÍA	 222
 ANEXOS	 232

ÍNDICE DE CUADROS

<i>Cuadro 1. 1 Marco conceptual propuesto por Cherly Gray sobre los factores relevantes para la utilización de las evaluaciones</i>	<i>32</i>
<i>Cuadro 1. 2 Mecanismos que pueden mediar la influencia de la evaluación, Mark y Henry, 2003.....</i>	<i>51</i>
<i>Cuadro 1. 3 Retos en la implementación de los resultados de las evaluaciones</i>	<i>58</i>
<i>Cuadro 2.1 Evaluaciones de Sedesol con valoración de GESOC, 2007-2014.....</i>	<i>70</i>
<i>Cuadro 2.2 Programas federales de Sedesol considerado en el análisis de evaluaciones.....</i>	<i>71</i>
<i>Cuadro 2.3 Casos en que un mismo evaluador realiza más de 50 por ciento de las valoraciones de un programa presupuestario</i>	<i>75</i>
<i>Cuadro 2.4 Coordinadores de evaluaciones que participaron con más de una institución para valorar los programas de Sedesol</i>	<i>76</i>
<i>Cuadro 2.5 Costo total de las evaluaciones de Sedesol según tipo de evaluación, 2007-2014</i>	<i>78</i>
<i>Cuadro 2.6 Origen del presupuesto de las evaluaciones por año, según tipo de evaluación.....</i>	<i>80</i>
<i>Cuadro 3. 1 Criterios para la realización de las Evaluaciones de Diseño, según los Términos de Referencia</i>	<i>102</i>
<i>Cuadro 3. 2 Temas considerados y número de preguntas en los términos de referencia para la elaboración de las evaluaciones de diseño.....</i>	<i>104</i>
<i>Cuadro 3. 3 Apartados considerados en los términos de referencia para las EEDS.....</i>	<i>106</i>
<i>Cuadro 3. 4 Factores endógenos analizados</i>	<i>110</i>
<i>Cuadro 3. 5 Valores medio, mínimo y máximo de las variables.....</i>	<i>113</i>
<i>Cuadro 3. 6 Matriz de correlación de las variables principales</i>	<i>115</i>
<i>Cuadro 3. 7 Análisis de regresión bivariada para factores endógenos.....</i>	<i>116</i>
<i>Cuadro 3. 8 Análisis de regresión multivariada para factores endógenos</i>	<i>119</i>
<i>Cuadro 3. 9 Factores exógenos o institucionales analizados.....</i>	<i>120</i>
<i>Cuadro 3. 10 Análisis.....</i>	<i>126</i>
<i>Cuadro 3. 11 Matriz de Correlación.....</i>	<i>127</i>
<i>Cuadro 3. 12 Análisis de regresión bivariada para explicar la utilización de las evaluaciones a partir de los arreglos institucionales</i>	<i>128</i>
<i>Cuadro 3. 13 Análisis de regresión multivariada para factores institucionales</i>	<i>132</i>

<i>Cuadro 4. 1 Recomendaciones de diseño emitidas y atendidas.....</i>	<i>142</i>
<i>Cuadro 4. 2 Valores medio, mínimo y máximo de las variables.....</i>	<i>162</i>
<i>Cuadro 4. 3 Matriz de correlación de las variables principales</i>	<i>164</i>
<i>Cuadro 4. 4 Análisis de regresión bivariada respecto a la valoración del INDEP</i>	<i>165</i>
<i>Cuadro 4. 5 El efecto de las EED en la valoración del INDEP</i>	<i>168</i>
<i>Cuadro 5. 1 Evaluaciones realizadas al PDZP</i>	<i>176</i>
<i>Cuadro 5. 2 Evaluaciones realizadas al Programa 3X1</i>	<i>176</i>
<i>Cuadro 5. 3 Síntesis de los hallazgos generales sobre los estudios de caso</i>	<i>177</i>
<i>Cuadro 5. 4 Evaluaciones analizadas del Programa para el Desarrollo de Zonas Prioritarias</i> <i>.....</i>	<i>179</i>
<i>Cuadro 5. 5 Evaluaciones analizadas del Programa 3X1 para Migrantes</i>	<i>180</i>
<i>Cuadro 5. 6 Recomendaciones y ASM derivados de las EED del PDZP, 2010 y 2014.....</i>	<i>184</i>
<i>Cuadro 5. 7 Recomendaciones y ASM derivados de las EED del Programa 3X1, 2008 y 2014</i> <i>.....</i>	<i>187</i>
<i>Cuadro 5. 8 Síntesis del efecto de la evaluación en los estudios de caso.....</i>	<i>193</i>
<i>Cuadro 5. 9 Recomendaciones derivadas de las ED implementadas en las ROP.....</i>	<i>195</i>

ÍNDICE DE GRÁFICOS

<i>Gráfico 1. Porcentaje de población en México en condiciones de pobreza y pobreza extrema, 2008-2014.....</i>	<i>6</i>
<i>Gráfico 2. Evolución en el desempeño de los Programas Federales 2008-2013.....</i>	<i>7</i>
<i>Gráfico 3. Grado de satisfacción con la democracia en México 2013.....</i>	<i>12</i>
<i>Gráfico 4. Indicadores de gobernanza en México, 2003-2013.....</i>	<i>16</i>
<i>Gráfico 2. 1 Porcentaje de evaluaciones realizadas por tipo de organismo evaluador.....</i>	<i>74</i>
<i>Gráfico 2. 2 Organismos evaluadores con mayor participación en la valoración de los programas federales de Sedesol</i>	<i>74</i>
<i>Gráfico 2. 3 Costo promedio de las evaluaciones realizadas según tipo de evaluación y procedencia del financiamiento, 2007 - 2014.....</i>	<i>77</i>
<i>Gráfico 2. 4 Proporción del presupuesto de los programas que representa el costo de las evaluaciones según tipo de evaluación, 2007 – 2014.....</i>	<i>79</i>
<i>Gráfico 2. 5 Gasto total y promedio de las evaluaciones según procedencia del presupuesto</i>	<i>81</i>
<i>Gráfico 2. 6 Porcentaje de recomendaciones que se convierten en ASM por tipo de evaluación.....</i>	<i>87</i>
<i>Gráfico 2. 7 Porcentaje de ASM atendidos por tipo de evaluación.....</i>	<i>87</i>
<i>Gráfico 2. 8 Porcentaje de recomendaciones que se convierten en ASM según tipo de evaluación y tema</i>	<i>87</i>
<i>Gráfico 2. 9 Porcentaje de ASM atendidos según tipo de evaluación y tema</i>	<i>87</i>
<i>Gráfico 2. 10 Porcentaje de recomendaciones que se convierten en ASM según el tema que aborda el programa</i>	<i>88</i>
<i>Gráfico 2. 11 Porcentaje de ASM atendidos según</i>	<i>88</i>
<i>Gráfico 2. 12 Porcentaje de recomendaciones que se convierten en ASM.....</i>	<i>89</i>
<i>Gráfico 2. 13 Porcentaje de ASM atendidos por programa</i>	<i>89</i>
<i>Gráfico 4. 1 Porcentaje de evaluaciones que atendieron las recomendaciones orientadas a las ROP (porcentaje)</i>	<i>143</i>
<i>Gráfico 4. 2 Temas que abordan las recomendaciones de diseño atendidas, 2007 y 2014 (porcentaje).....</i>	<i>145</i>
<i>Gráfico 4. 3 Número de casos donde se observaron cambios en las ROP por tema,.....</i>	<i>147</i>

<i>Gráfico 4. 4 Cambios registrado en las reglas de operación (número de variables modificadas por año seleccionado)</i>	<i>148</i>
<i>Gráfico 4. 5 Cambios registrado en variables de las reglas de operación a partir de las evaluaciones (número de variables modificadas por año seleccionado)</i>	<i>148</i>
<i>Gráfico 4. 6 Variación del valor del Índice de Calidad del Diseño (ICADI) 2007-2011.....</i>	<i>151</i>
<i>Gráfico 4. 7 Comparación de los Índices de desempeño de los programas de Sedesol respecto al resto de los programas federales evaluados, 2008-2014</i>	<i>156</i>
<i>Gráfico 4. 8 Variación del INDEP en cada programa de Sedesol evaluado, 2008-2014</i>	<i>157</i>
 <i>Gráfico 5. 1 El efecto de las recomendaciones de diseño en el ICADI</i>	 <i>197</i>
<i>Gráfico 5. 2 El efecto de las evaluaciones en el INDEP.....</i>	<i>199</i>

ÍNDICE DE ILUSTRACIONES

<i>Ilustración 1. 1 Posibles escenarios para el uso de los hallazgos de la evaluación</i>	<i>28</i>
<i>Ilustración 1. 2 Factores determinantes en la utilización de las evaluaciones de intervenciones gubernamentales, según su cercanía al proceso evaluativo..</i>	<i>35</i>
<i>Ilustración 1. 3. Marco de análisis de Karen Kirkhart sobre la Influencia (2000).....</i>	<i>49</i>
<i>Ilustración 1. 4 Nuevo modelo sobre la Influencia propuesto por Alkin y Taut (2003), adaptando el realizado por Kirkhart (2000).....</i>	<i>50</i>
<i>Ilustración 1. 5 Factores determinantes en la implementación de los resultados de las evaluaciones</i>	<i>58</i>
<i>Ilustración 2. 1 Posibles escenarios sobre la utilización de la práctica evaluativa.....</i>	<i>67</i>
<i>Ilustración 2. 2 Diferentes etapas para la utilización de las evaluaciones</i>	<i>82</i>
<i>Ilustración 4. 1 Recomendaciones de diseño que fueron atendidas para mejorar las ROP, 2007 y 2014</i>	<i>142</i>
<i>Ilustración 5. 1 Hallazgos sobre la utilización de la evaluación y la capacidad explicativa de los factores endógenos y exógenos.....</i>	<i>182</i>
<i>Ilustración 5. 2 Hallazgos sobre el efecto de la práctica evaluativa en el desempeño de los programas: la valoración del INDEP y la capacidad explicativa de la implementación de los ASM</i>	<i>192</i>
<i>Ilustración 6. 1 Posturas de diferentes actores sobre el tipo de aprovechamiento que debe hacerse de los hallazgos derivados de las evaluaciones.....</i>	<i>208</i>

SIGLAS Y ACRÓNIMOS

DEFINICIONES BÁSICAS Y GLOSARIO DE

TÉRMINOS

SIGLAS Y ACRÓNIMOS

3x1	Programa 3x1 para Migrantes
70 y Más	Programa 70 y Más, ahora Pensión para Adultos Mayores
Alternativas en Economía	Alternativas en Economía, S.A. De C.V.
APF	Administración Pública Federal
ASF	Auditoría Superior de la Federación
ASM	Aspectos Susceptibles de Mejora
BetaKorosi Consultores	BetaKorosi Consultoría S.C.
BID	Banco Interamericano de Desarrollo
C-230 Consultores	C-230 Consultores, S. C.
CEESS	Centro de Estudios Económicos y Sociales en Salud, Hospital Infantil de México Federico Gómez
CEPNA	Centro de Estudios para un Proyecto de Nacional Alternativo S.C.
Chapingo	Universidad Autónoma Chapingo
CIAD	Centro de Investigación en Alimentación y Desarrollo, A.
CIDE	Centro de Investigación y Docencia Económica, A.C.
CIDEA	Consultores en Innovación Desarrollo y Estrategia Aplicada S.C.
CIESAS	Centro de Investigaciones y Estudios Superiores en Antropología Social
CISS	Conferencia Interamericana de Seguridad Social
CLAD	Centro Latinoamericano de Administración para el Desarrollo
CNPDHO	Coordinación Nacional del Programa de Desarrollo Humano Oportunidades
COLEF	El Colegio de la Frontera Norte, A.C.
Colegio de Economistas	Colegio Nacional de Economistas A.C.

Colegio de Postgraduados	Colegio de Postgraduados
Colegio Mexiquense	El Colegio Mexiquense, A.C.
COLMEX	El Colegio de México, A.C.
Coneval	Consejo Nacional de Evaluación de la Política de Desarrollo Social
DG	Director General
DGA	Director General Adjunto
Diconsa	Diconsa, S.A. de C.V.
ED	Evaluaciones de Diseño
EED	Evaluaciones Específicas de Desempeño
Evalutare	Evalutare Expertos en Evaluación de Políticas Públicas S.A. de C.V.
FAO	Organización de las Naciones Unidas para la Agricultura y la Alimentación
Fonart	Fondo Nacional para el Fomento de las Artesanías
Fonhapo	Fondo Nacional de Habitaciones Populares
GESOC	Organización de sociedad civil para la Gestión Social y Cooperación, A.C
GpR	Gestión para Resultados
Hábitat	Programa Hábitat
ICADI	Índice de Calidad de Diseño
Imjuventud	Instituto Mexicano de la Juventud
INAP	Instituto Nacional de Administración Pública
INDEP	Índice de Desempeño de Programas Públicos Federales
Indesol	Instituto Nacional de Desarrollo Social
INEGI	Instituto Nacional de Estadística y Geografía
INSP	Instituto Nacional de Salud Pública
LDS	Ley de Desarrollo Social
LGEPF	Lineamientos Generales para la Evaluación de los Programas Federales

Liconsa	Liconsa, S.A. de C.V.
Lineamientos Generales	Lineamientos Generales para la Evaluación de los Programas Federales de la Administración Pública Federal, 2007
MCO	Método estadístico de Mínimos Cuadrados Ordinarios
Mecanismo de Seguimiento	Mecanismo para el Seguimiento a los Aspectos Susceptibles de Mejora
MIR	Matriz de Indicadores para Resultados
MML	Matriz de Marco Lógico
NGP	Nueva Gestión Pública
NIK Beta	NIK Beta, S.C.
OIC	Órgano Interno de Control
OSC	Organizaciones de la Sociedad Civil
PAAMZR	Programa de Atención a Adultos Mayores en Zonas Rurales
PAE	Programa Anual de Evaluación
PAIMEF	Programa de Apoyo a las Instancias de Mujeres en las Entidades Federales para Implementar y Ejecutar Programas de Prevención de la Violencia contra las Mujeres
PAJA	Programa de Atención a Jornaleros Agrícolas
PAL	Programa de Apoyo Alimentario
PAR	Programa de Abasto Rural a cargo de Diconsa, S.A. de C.V.
PASCV	Programa de Ahorro, Subsidio y Crédito para la Vivienda “Tu Casa”
PASL	Programa de Abasto Social de Leche a cargo de Liconsa, S.A. de C.V.
PASPRAH	Programa de Apoyo a las Vecindades en Condiciones de Pobreza Patrimonial para Regularizar Asentamientos Humanos Irregulares
PbR	Presupuesto basado en Resultados
PCS	Programa Coinversión Social
PDHO / PROSPERA	Programa de Desarrollo Humano Oportunidades / Prospera
PDL	Programa Desarrollo Local (Microrregiones)

PDZP	Programa para el Desarrollo de Zonas Prioritarias
PEI	Programa Estancias Infantiles para Apoyar a Madres Trabajadoras
PET	Programa Empleo Temporal
PNUD	Programa de las Naciones Unidas para el Desarrollo
POP	Programa Opciones Productivas
PP	Programa Presupuestario
PREP	Programa de Rescate de Espacios Públicos
Programa FONART	Programa del Fondo Nacional de Fomento a las Artesanías
PSPJ	Subsidio a Programas para Jóvenes
PSVJF	Programa Seguro de Vida para Jefas de Familia
PVR	Programa Vivienda Rural
Sedesol	Secretaría de Desarrollo Social
SFP	Secretaría de la Función Pública
SHCP	Secretaría de Hacienda y Crédito Público
Sisee	Sistema de Seguimiento a Evaluaciones Externas
SMYE	Sistema de Monitoreo y Evaluación
UAEM	Universidad Autónoma del Estado de México
UAM	Universidad Autónoma Metropolitana
UAT	Universidad Autónoma de Tamaulipas
UED	Unidad de Evaluación del Desempeño de la SHCP
UIEM	Universidad Intercultural del Estado de México
UNAM	Universidad Nacional Autónoma de México
ZAP	Zonas de Atención Prioritarias

DEFINICIONES BÁSICAS Y GLOSARIO DE TÉRMINOS

Evaluación de programas públicos: análisis sistemático y objetivo de programas y acciones federales, que tiene como finalidad determinar la pertinencia y el logro de sus objetivos y metas, así como su eficiencia, eficacia, calidad, resultados, impacto y sostenibilidad (SHCP).

Evaluación externa: aquella que se realiza a través de personas físicas y morales especializadas y con experiencia probada en la materia que corresponda evaluar, que cumplan con los requisitos de independencia, imparcialidad, transparencia y los demás que se establezcan en las disposiciones aplicables (Coneval, SHCP y SFP, 2015).

Evaluación de Consistencia y Resultados: aquellas que analizan sistemáticamente el diseño y desempeño global de los programas federales, para mejorar su gestión y medir el logro de sus resultados con base en la matriz de indicadores (SHCP, SFP y Coneval, 2007).

Evaluaciones de diseño: aquellas derivadas de las Evaluaciones de Consistencia y Resultados, a la que se refiere la fracción I del numeral Décimo Octavo de los Lineamientos Generales para la evaluación de los Programas Federales de la Administración Pública Federal (SHCP et al., 2007).

Evaluaciones de desempeño: las Evaluaciones Específicas de Desempeño (EED) son aquellas no comprendidas en los lineamientos y que se realizan mediante trabajo de gabinete o de campo con documentos e indicadores establecidos por el Coneval para tal efecto (SHCP et al., 2007 y Coneval, SHCP y SFP, 2015).

Aspectos Susceptibles de Mejora: hallazgos, debilidades, oportunidades y amenazas que se identifican en la evaluación externa, las cuales pueden ser atendidas para la mejora de los programas con base en las recomendaciones y sugerencias señaladas por el evaluador externo a fin de mejorar los programas presupuestarios (SHCP)¹.

¹ Una vez realizada la evaluación, se procede a la selección de los Aspectos Susceptibles de Mejora (ASM) conforme a criterios de claridad, relevancia, justificación y factibilidad. Información consultada en: http://www.shcp.gob.mx/EGRESOS/sitio_pbr/evaluacion/Paginas/ASM.aspx.

Cada dependencia selecciona y clasifica sus ASM, entendiéndose éstos como las acciones con las cuales se dará atención a los hallazgos, debilidades, oportunidades y amenazas, y cuya finalidad es la mejora del programa. En: <http://www.funcionpublica.gob.mx/index.php/ua/ssfp/uegdg/pae.html>

Coherencia interna: en la que se analiza y juzga, por una parte, la compatibilidad y las sinergias entre objetivos, y por otra parte, la asociación lógica entre los medios y los objetivos que se pretenden alcanzar. Es necesario considerar tanto los objetivos explícitos como los implícitos dado que la pretensión es que sean alcanzados ambos, muchas veces más los segundos que los primeros. A través de la evaluación de la coherencia interna se analiza la ordenación de los niveles en los que se establece la estrategia y su lógica en virtud de los “modelos de repercusión” (Osuna y Márquez, 2000).

Actividades desarrolladas por los programas: iniciativas planeadas y llevadas a cabo por las dependencias gubernamentales responsables de la operación del programa para cumplir con los objetivos definidos y para atender los problemas sociales que justifican su existencia.

Desempeño de programas: la rentabilidad social de cada programa y su capacidad para cumplir realmente los objetivos para los que fue creado, entregando resultados relevantes a la ciudadanía (GESOC, 2009).

Efectos de la evaluación: Cambios en el diseño o en el desempeño de los programas públicos, derivados de las recomendaciones de las evaluaciones, que impactan en su capacidad para lograr los objetivos.

RESUMEN

RESUMEN

La práctica evaluativa ha sido adoptada por numerosos gobiernos alrededor del mundo, desde la década de 1990, con el propósito de mejorar la gestión pública a través de un enfoque de Gestión para Resultados. Sin embargo, poco se sabe sobre el efecto real que las evaluaciones tienen en el mejoramiento de los programas. En este sentido, a diez años de la institucionalización de la práctica evaluativa en México, el objetivo de esta tesis es explorar qué tanto se han utilizado las evaluaciones realizadas a la Secretaría de Desarrollo Social (Sedesol) entre 2007 y 2014, así como su efecto tanto en el diseño como en el desempeño de los programas valorados.

El propósito de esta investigación es ofrecer evidencia empírica en torno al aprovechamiento real de las evaluaciones de Sedesol y evidenciar que los ajustes realizados, si bien representan un avance importante, ha tenido un alcance limitado en la gestión pública. El estudio del efecto de las evaluaciones en los programas públicos cobra especial importancia en el contexto latinoamericano, donde los países han enfrentado crisis institucionales que coartaron su capacidad de formular políticas de Estado que respondan a las demandas ciudadanas; esto debido a que si bien se han logrado avances en materia democrática, en la mayoría de estos países no se han generado las condiciones para producir un desarrollo económico y social sostenible (Rojas Aravena, 2004). Además, el énfasis de esta investigación en los programas sociales es de gran pertinencia a razón de los altos niveles de desigualdad y pobreza que predominan en América Latina; mismos que poco han evolucionado a partir de la instauración de gobiernos democráticos.

En esta investigación, las principales fuentes de información que empleé para analizar la utilización y el efecto de las evaluaciones fueron: la revisión de documentos oficiales; el análisis de 111 reportes de evaluaciones externas (36 de diseño y 75 de desempeño) emitidas entre 2007 y 2014; los informes de seguimiento a Aspectos Susceptibles de Mejora (ASM); los índices que estima la organización de la sociedad civil Gestión Social y Cooperación, A.C. (GESOC), mediante los cuales estima la calidad de su diseño (ICADI) y la capacidad de los programas para lograr sus objetivos (INDEP); y, por último, entrevistas a actores claves. La información obtenida de las fuentes mencionadas fue examinada y sistematizada para crear bases de datos que me permitieron realizar análisis cuantitativos y cualitativos: por un lado, un análisis cuantitativo para estimar qué tanto se utilizaron las evaluaciones y su relevancia en

la mejora del diseño y el desempeño de los programas, a través de métodos estadísticos de regresión lineal multivariada. Por otro lado, un análisis cualitativo que me permitió explicar los hallazgos obtenidos con los modelos estadísticos.

El documento consta de seis capítulos. En el primero, abordo brevemente la literatura disponible sobre el aprovechamiento de las evaluaciones y sus efectos. Cabe mencionar que los estudios que se han realizado desde la década de los 1970 para entender la utilización de las evaluaciones se han centrado, principalmente, en identificar cuáles son los factores determinantes (propios o ajenos al proceso evaluativo), sin que exista en la literatura un consenso sobre cómo éstos deben ser estimados o la relevancia de tienen cada uno de ellos. Además, que desde inicios del siglo XXI, existe un creciente interés por conocer cuál es el efecto de las evaluaciones; no obstante, es un campo de investigación todavía en desarrollo que ha generado muy pocas contribuciones empíricas.

En el segundo capítulo, presento un panorama inicial sobre las evaluaciones realizadas a los programas de Sedesol y su utilización. Para ello, exploré las 111 evaluaciones externas realizadas a los programas de Sedesol bajo la coordinación del Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval). Los resultados señalan que el aprovechamiento de las evaluaciones ha sido considerable, más de 50 por ciento de las recomendaciones identificadas para el periodo fueron formalizadas como ASM por las unidades responsables de los programas, y aproximadamente 65 por ciento de esos ASM fueron efectivamente implementados. Sin embargo, esta tendencia no fue constante a lo largo del periodo, se registró una disminución gradual sostenida tanto en la formalización como en la implementación de las recomendaciones.

En el tercer capítulo, analizo en qué medida algunas características de la evaluación (factores endógenos) y la independencia del proceso evaluativo (factores exógenos o institucionales) ayudan a explicar que sus resultados sean utilizados, aplicando dos análisis de regresión multivariada de mínimos cuadrados ordinarios (MCO). A partir de este análisis, encontré que la utilización de las evaluaciones puede explicarse parcialmente a través de factores como: la difusión de los informes (resultados en Google), la intensidad de las evaluaciones (número de evaluaciones acumuladas), el costo y la distancia entre evaluador y evaluado.

En el cuarto capítulo, exploro el efecto de las evaluaciones de diseño y desempeño identificando los cambios que las recomendaciones atendidas generaron en las Reglas de Operación de los Programas (ROP) y en la capacidad de estos para alcanzar sus objetivos (INDEP). Por un lado, mediante un análisis comparativo, encontré que aproximadamente 37 por ciento de las recomendaciones de diseño aplicables a las ROP de los programas fueron implementadas. Por otro lado, con un análisis de regresión multivariado, descubrí que la implementación de las recomendaciones derivadas de Evaluaciones Específicas de Desempeño (EED) ayuda a explicar el desempeño posterior de los programas pero en una proporción modesta.

En el quinto capítulo, a través de un estudio de caso, analizo la experiencia de dos programas sociales orientados a combatir la pobreza: Programa para el Desarrollo de Zonas Prioritarias (PDZP) y 3X1 para Migrantes (3X1). A través de estos, identifiqué que una participación activa por parte de la unidad de evaluación de Sedesol es determinante para la utilización de las evaluaciones, específicamente para la formalización de ASM. Además, que las mejoras que registran los programas en la calidad de su diseño (estimado a través del ICADI) no son atribuibles a los cambios realizados en ROP al atender las recomendaciones derivadas de las Evaluaciones de Diseño (ED). En el sexto y último capítulo, presento las conclusiones de la presente investigación y una breve reflexión sobre los alcances de la práctica evaluativa.

ABSTRACT

ABSTRACT

The evaluation practice has been adopted by numerous governments around the world, since the 1990s, with the aim of improving public management through a Managing for Result approach. However, little is known about the real impact of evaluations on program improvement. In this sense, ten years after the institutionalization of the evaluation practice in Mexico, the objective of this thesis is to explore how much the evaluations made to the Ministry of Social Development (Sedesol, for its acronym in spanish) have been made between 2007 and 2014, as well as its effect both in the design and in the performance of the evaluated programs.

The purpose of this research is to provide empirical evidence on the actual utilization of Sedesol evaluations and to show that the adjustments made, although representing a significant advance, have had limited scope in public management. The study of the effect of evaluations on public programs is especially important in the Latin American context, where countries have faced institutional crises that have restricted their ability to formulate State policies that respond to citizen demands. This is because, although progress has been made in democratic matters, the conditions for producing sustainable economic and social development have not been generated in most of these countries (Rojas Aravena, 2004). In addition, the emphasis of this research on social programs is of great relevance because of the high levels of inequality and poverty prevailing in Latin America, which have hardly evolved since the establishment of democratic governments.

In this research, the main sources of information used to analyze the utilization and effect of the evaluations were: the revision of official documents, the analysis of 111 reports of external evaluations (36 of design and 75 of performance) issued between 2007 and 2014, follow-up reports on Susceptible Aspects of Improvement (ASM, for its acronym in spanish), the indexes estimated by GESOC (ICADI to estimate the quality of the design and INDEP for the capacity of the programs to achieve their objective), and interviews with key stakeholders. The information obtained from the mentioned sources was examined and systematized to create databases that allowed me to carry out quantitative and qualitative analyzes: on the one hand, a quantitative analysis to estimate how much the evaluations were utilized and their relevance in the improvement of the design and the Performance of the programs, through

statistical methods of multivariate linear regression. On the other hand, a qualitative analysis that allowed me to explain the findings obtained with the statistical models.

The document consists of six chapters. In the first, I review the available literature on the application of evaluations and their effects. It should be mentioned that the studies that have been carried out since the 1970s to understand the utilization of evaluations have focused mainly on identifying which are the determinants (own or outside the evaluative process), without existing in the literature a consensus on how these should be estimated or the relevance of each of them. Moreover, since the beginning of the 21st century, there has been a growing interest in knowing the effect of evaluations. However, it is a field of research still under development that has generated very few empirical contributions.

In the second chapter, I present an initial overview of the evaluations of Sedesol programs and their utilization. To do this, I explored the 111 external evaluations made to the Sedesol programs under the coordination of the National Council for Evaluation of Social Development Policy (Coneval, for its acronym in spanish). The results indicate that the utilization of the evaluations has been considerable, more than 50 percent of the recommendations identified for the period were formalized as ASM by the units responsible for the programs, and approximately 65 percent of those ASM were effectively implemented. However, this trend was not constant throughout the period, there was a gradual decrease in both the formalization and the implementation of the recommendations.

In the third chapter, I analyze to what extent some characteristics of the evaluation (endogenous factors) and the independence of the evaluative process (exogenous or institutional factors) help to explain that their results are applied, using two ordinary least squares multivariate regression analyzes. Based on this analysis, I found that the utilization of evaluations can be partially explained by factors such as: dissemination of reports (results provided by Google), intensity of evaluations (number of evaluations accumulated), cost and distance between evaluator and evaluated.

In the fourth chapter, I explore the effect of design and performance evaluations by identifying the changes that the recommendations addressed in the ROP and their ability to achieve their objectives (INDEP, for its acronym in spanish). On the one hand, through comparative analysis, I found that approximately 37 percent of the design recommendations

applicable to program ROP were implemented. On the other hand, with a multivariate regression analysis, I found that the implementation of the recommendations derived from Specific Performance Evaluations (EEE, for its acronym in spanish) helps to explain the later performance of the programs but in a very modest proportion.

In the fifth chapter, through a case study, I analyze the experience of two social programs aimed at combating poverty: Program for the Development of Priority Zones (PDZP) and 3X1 for Migrants (3X1). Through these, I identified that an active participation on the part of the evaluation unit of Sedesol is determinant for the utilization of the evaluations, specifically for the formalization of ASM. In addition, improvements in the quality of their design (estimated through ICADI) are not attributable to the changes made to ROP by complying with the recommendations derived from the Design Evaluations (ED, for its acronym in spanish). In the sixth and final chapter, I present the conclusions of the present research and a brief reflection on the scope of evaluative practice.

INTRODUCCIÓN

INTRODUCCIÓN

El objeto de estudio de esta tesis es la práctica evaluativa como mecanismo para promover mejoras en los programas públicos. Este trabajo fue motivado por el auge que ha tenido en las últimas décadas la Gestión para Resultados (GpR), a partir del cual la evaluación ha sido adoptada por numerosos gobiernos alrededor del mundo para responder a su necesidad de incrementar la capacidad gubernamental y la gestión pública, sin que exista evidencia suficiente sobre los beneficios reales que esta práctica genera.

En el caso de México, a razón del papel estratégico que se le ha otorgado a la evaluación sistemática de programas para atender los problemas de eficiencia y eficacia gubernamental, consideré necesario analizar los aportes que esta ha realizando para determinar si la estrategia ha logrando su cometido o si es necesario someterla a un replanteamiento. En este sentido, el objetivo de este trabajo es analizar la utilización y el efecto de las evaluaciones de diseño y desempeño realizadas a los programas de la Secretaría de Desarrollo Social (Sedesol) entre 2007 y 2014, y determinar si la implementación de las recomendaciones efectivamente les ha ayudado a los programas a incrementar la capacidad para alcanzar sus objetivos.

En 2004, el gobierno mexicano dio el primer paso hacia la construcción un Sistema de Evaluación del Desempeño con la publicación de la Ley de Desarrollo Social (LDS), donde se establecen estándares de Monitoreo y Evaluación (MyE) internacionalmente aceptados para la evaluación de los programas públicos. Posteriormente, en 2006, estos esfuerzos se institucionalizan con la creación del Consejo Nacional de Evaluación de la Política Social (Coneval) que tiene la misión de medir los resultados nacionales en materia de reducción de la pobreza y coordinar la evaluación de los programas sociales en el gobierno federal (Fernando Castro, et al., 2009; Banco Mundial, 2009). A partir de 2007, el Coneval junto con la Secretaría de Hacienda y Crédito Público (SHCP) ha liderado a los programas sociales en el impulso de la GpR con el propósito de mejorar su desempeño y hacer más eficiente en el gasto público.

Hoy en día, a diez años de que se institucionalizó la evaluación sistemática de programas sociales en México, no queda claro que tanto ésta práctica ha favorecido mejoras

significativas en el desempeño de los programas orientados a combatir la pobreza. Sin embargo, los datos disponibles sugieren que las mejoras han sido modestas y su efecto imperceptible. Por un lado, el índice de desempeño que estima anualmente la organización de la sociedad civil para la Gestión Social y Cooperación, A.C. (GESOC), señala que entre 2008 y 2014, el desempeño de los programas evaluados varió poco de una calificación promedio de 5.6 a 6.2, en un rango de uno a diez. Por otro lado, se registra un suceso todavía más preocupante, un crecimiento en los niveles de pobreza a nivel nacional que alcanzó en 2014, el 55.7 por ciento de la población.

En este sentido, el riesgo que se corre si no se identifican y atienden oportunamente los retos que enfrenta la práctica evaluativa en el caso mexicano, es que se mantenga la misma tendencia mostrada hasta ahora, donde se destina una importante cantidad de recursos económicos y humanos para cumplir este costoso requisito administrativo que ha servido principalmente para evidenciar los problemas existentes pero sin lograr un impacto significativo en el desempeño de los programas valorados y sin contribuir al mejoramiento de las condiciones de vida de la población.

Los estudios sobre el aprovechamiento de las evaluaciones datan de la década de los 1970s, cuando surge el interés por explicar a través del análisis de diversos factores (propios o ajenos a la práctica) qué determina que los resultados de las evaluaciones sean utilizados. Desde inicios del siglo XXI, las investigaciones se han orientado lentamente hacia el estudio del impacto que generan las evaluaciones utilizadas; no obstante, hoy en día todavía existen muy pocas contribuciones empíricas sobre los mecanismos que pudiesen describir cómo tienen lugar el efecto de la evaluación (Herbert, 2014). A través de esta investigación busco aportar evidencia empírica referente a la implementación de las recomendaciones derivadas de las evaluaciones y su efecto en los programas, haciendo hincapié en cómo dichas recomendaciones impactan de manera inmediata en el diseño y, de manera no inmediata, en su desempeño.

Los estudios sobre el efecto de la evaluación en los programas cobra especial importancia en el contexto latinoamericano, donde los países se han visto envueltos en crisis institucionales que han coartado su capacidad de formular políticas de Estado que respondan a las demandas ciudadanas; si bien se han logrado avances en materia democrática, en la mayoría de estos países no se han generado las condiciones para producir un desarrollo

económico y social sostenible (Rojas Aravena, 2004). Asimismo, el énfasis en el efecto sobre los programas sociales es de gran pertinencia a razón de los altos niveles de desigualdad y pobreza que predominan en América Latina; mismos que poco han evolucionado a partir de la instauración de gobiernos democráticos.

En esta investigación, las fuentes de información que empleé para analizar la utilización y el efecto de las evaluaciones fueron: i) la revisión de documentos oficiales, ii) el análisis de los reportes emitidos por evaluadores externos que colaboraron con Sedesol entre 2007 y 2014, iii) los índices que estima anualmente GESOC (ICADI e INDEP), y iv) entrevistas a actores claves. La información obtenida de las fuentes mencionadas fue sistematizada para crear una base de datos, y posteriormente, analizada a través de métodos estadísticos con el propósito de explicar la relevancia de las evaluaciones en la mejora del diseño y el desempeño de los programas; para ello, se empleó análisis cualitativo y de regresión lineal multivariada.

El documento consta de seis capítulos. En el primero, abordo brevemente la literatura disponible sobre el aprovechamiento de las evaluaciones y sus efectos. En el segundo, presento un panorama inicial sobre las evaluaciones realizadas a los programas de Sedesol y su utilización. En el tercero, analizo en qué medida algunas características de la evaluación y la independencia del proceso evaluativo ayudan a explicar que sus resultados sean utilizados. En el cuarto, exploro el efecto de las recomendaciones de diseño y desempeño implementadas a partir de cambios en las Reglas de Operación de los Programas y en la capacidad para alcanzar sus objetivos (INDEP). En el quinto, a través de un estudio de caso, analizo la experiencia de dos programas sociales orientados al combate a la pobreza: Programa para el Desarrollo de Zonas Prioritarias (PDZP) y 3X1 para Migrantes (3X1). En el sexto y último capítulo, presento las conclusiones de la presente investigación y una breve reflexión sobre los alcances de la práctica evaluativa.

Los alcances de esta investigación son descriptivos y explicativos; descriptivos en la medida en que los resultados de ese trabajo permitirán conocer los efectos que han tenido las evaluaciones sobre el diseño y el desempeño de los programas; y explicativo, a razón de que ayudará a entender la relación entre la atención de las recomendaciones y las mejoras que han sido experimentadas, o no, por los programas evaluados. Esta tesis no pretendió ser una meta-evaluación o un meta-análisis de la práctica de evaluación en México.

JUSTIFICACIÓN PRÁCTICA

La Gestión para Resultados (GpR) es una estrategia integral que brinda un marco de referencia que facilita a las organizaciones públicas la dirección efectiva en su proceso de creación de resultados, a fin de optimizarlo, asegurando la máxima eficacia y eficiencia de su desempeño, la consecución de los objetivos de gobierno y la mejora continua de sus instituciones (YE y CLAD, 2007: en García López y García Moreno, 2010). Para ello, toma en cuenta los distintos elementos del ciclo de gestión: planificación, presupuesto, gestión financiera, gestión de proyectos, monitoreo y evaluación (García López y García Moreno, 2010).

En este marco de referencia, de acuerdo con el PNUD², la evaluación es una capacidad funcional necesaria para formular, implementar y revisar políticas y programas. Así como para “lograr que las cosas se hagan”, ya que permite medir los resultados y obtener comentarios para ajustar las políticas, codificar las lecciones aprendidas y promover el aprendizaje, así como para garantizar la rendición de cuentas a todos los actores involucrados (PNUD, 2008).

En las últimas décadas el gobierno mexicano ha impulsado diversas estrategias para introducir la evaluación en la Administración Pública Federal (APF), en un inicio promoviendo solamente la eficiencia del gasto público y más recientemente también la obtención de resultados. En los 1990s, la Secretaría de Desarrollo Social (Sedesol) fue pionera en los esfuerzos por implementar un sistema de Monitoreo y Evaluación (MyE) sectorial y en la utilización progresiva de información de las evaluaciones en la toma de decisiones de política social, lo que produjo una importante influencia sobre otras secretarías y sobre el gobierno federal (Fernando Castro et al., 2009, 4).

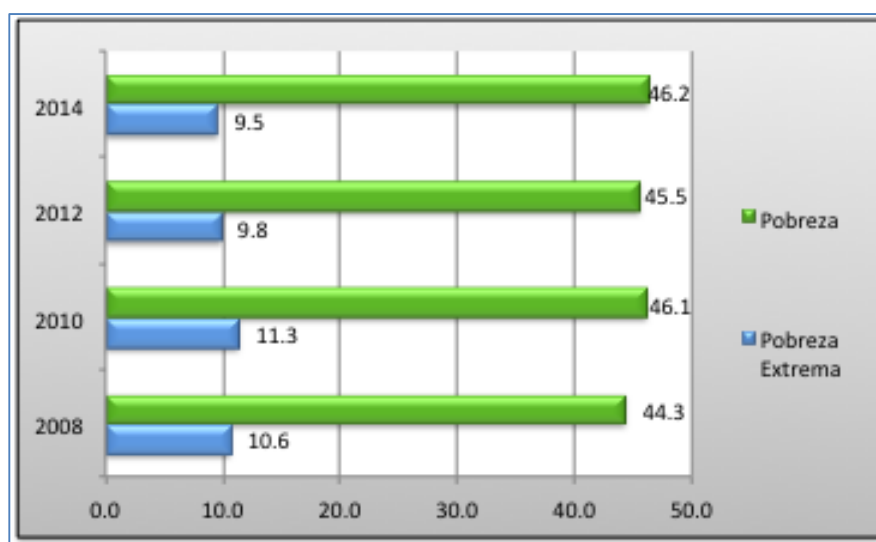
En 2004, el gobierno mexicano institucionalizó la evaluación de programas públicos a nivel nacional con el objetivo de hacer más eficiente el gasto público y mejorar el desempeño gubernamental. Desde entonces se le ha otorgado un papel estratégico a la evaluación de

² Las cinco capacidades funcionales identificadas por el PNUD son: i) capacidad para hacer que los actores involucrados participen; ii) capacidad para diagnosticar una situación y definir una visión y un mandato; iii) capacidad para formular políticas y estrategias; iv) capacidad para presupuestar, gestionar e implementar; y v) capacidad para evaluar.

programas y cada año se asigna un presupuesto mayor para este propósito (en 2015 el monto ascendió a 490 millones de pesos, un incremento del 35.8 por ciento respecto al año anterior, a precios constantes); por otra parte, cabe mencionar que la Secretaría de Hacienda y Crédito Público (SHCP) anunció la eliminación o fusión de más del 22 por ciento de los programas presupuestarios para 2016, aludiendo para esta decisión el seguimiento a los resultados de las evaluaciones. No obstante lo anterior, la evolución de esta práctica ha sido lenta, sus beneficios en la gestión pública son apenas perceptibles y no ha tenido hasta ahora una aplicación clara de un Presupuesto basado en Resultados (PbR).

A pesar de que han pasado casi diez años de la institucionalización de la evaluación, los programas sociales orientados a combatir la pobreza en México, así como la mayoría de los programas públicos, siguen siendo incapaces de alcanzar su fin último: mejorar las condiciones de vida de la población (*véase gráfica 1*). Entre 2008 y 2014, según datos del Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval), los niveles de pobreza en el país no han disminuido y más de la mitad de la población mexicana se encuentra en esta situación.

Gráfico 1. Porcentaje de población en México en condiciones de pobreza y pobreza extrema, 2008-2014

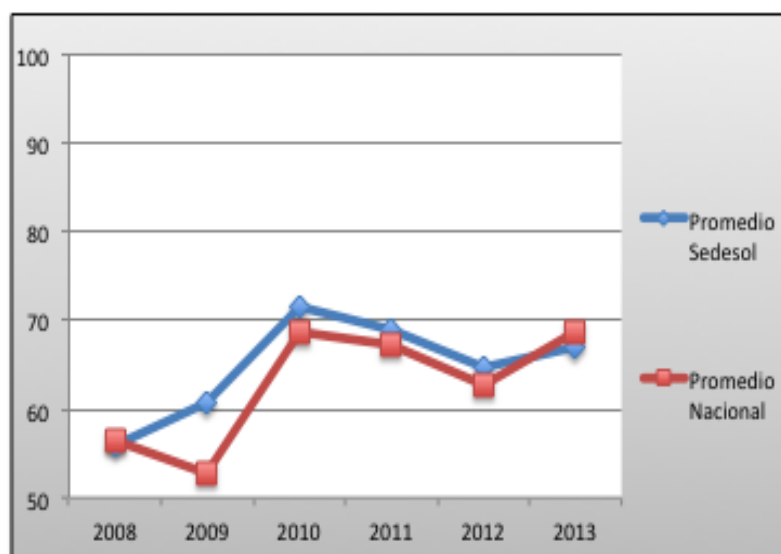


Fuente: Coneval.

De acuerdo con el Índice de Desempeño de Programas Públicos Federales (INDEP), desarrollado por la GESOC³ para valorar el logro de objetivos en los programas públicos evaluados. La calificación promedio del desempeño, entre 2008 y 2014, ascendió de 56 a 62 puntos en un rango de cero a cien (*véase gráfica 2*); no obstante, cabe destacar que esta mejoría en el desempeño de los programas se observa inconsistente a lo largo del tiempo, debido a que en algunos años se registraron crecimientos significativos seguidos de tendencias a la baja.

Los programas de la Sedesol que fueron evaluados también muestran una evolución irregular en su desempeño. Presentan variaciones que hasta cierto punto podrían ser atribuibles al costo de aprendizaje de la práctica de evaluación recientemente formalizada. No obstante, al margen de un significativo ascenso entre 2008 y 2010, la tendencia observada en años posteriores muestra retrocesos que crean incertidumbre sobre la capacidad de la evaluación para mejorar la gestión pública en el caso mexicano.

Gráfico 2. Evolución en el desempeño de los Programas Federales 2008-2013



Fuente: Índice de Desempeño de los Programas Federales GESOC 2009-2014.

Por tal motivo, se considera necesario analizar el efecto que la práctica de evaluación está generando en el desempeño de los programas sociales y su relación con la obtención mayores

³ La organización de sociedad civil para la Gestión Social y Cooperación, A.C. (GESOC), realiza anualmente estudios mediante los cuales valora la rentabilidad social de cada programa —entendida como su capacidad para cumplir con los objetivos para los que fueron creados, apenas muestran avances significativos en el cumplimiento de los objetivos y las metas de los programas.

resultados que permitan mejorar la situación de la población en condición de pobreza y extrema pobreza; esto a fin de determinar si la evaluación pública en México está siguiendo el camino correcto o si la estrategia requiere un replanteamiento. El riesgo de no abordar este tema radica en que la práctica de evaluación en México figure solo como un requisito administrativo intrascendente en la gestión de la APF.

JUSTIFICACIÓN TEÓRICA

En el ámbito de los gobiernos democráticos, la evaluación de programas públicos y el mejoramiento de su desempeño se relaciona con otros dos campos de investigación de gran relevancia social: i) la satisfacción con la democracia, y ii) la eficiencia directiva del gobierno.

En el marco de la democracia, los representantes públicos deben atender las demandas sociales a fin de garantizar el apoyo de la ciudadanía en los procesos electorales y así su permanencia en el poder. Los programas sociales son el principal vehículo por el cual los gobiernos atienden las necesidades básicas de la población, por lo que es de suma relevancia que éstos cuenten con un adecuado desempeño y cumplan con los objetivos establecidos.

La evaluación de programas es una importante herramienta que ayuda a medir el cumplimiento de los objetivos fijados y a generar procesos de retroalimentación que permiten mejorar el desempeño de los programas, así como su impacto social. Por lo tanto, si la práctica de evaluación logra mejores resultados y que las necesidades de la población sean efectivamente cubiertas, es de esperar que también produzca una mayor satisfacción de la ciudadanía con el gobierno.

Por otro lado, la influencia de los debates sobre la gobernanza, entendida como *una nueva manera de gobernar que introduce al ejercicio de la autoridad pública la participación de otros actores de la sociedad*, podrían favorecer el surgimiento de políticas y programas públicos más acertados, con un diseño e implementación propicios para satisfacer las demandas de la población. En el caso de los programas sociales de combate a la pobreza, que encaran una gran dificultad para cambiar la realidad de población vulnerable, éstos podrían beneficiarse de una gestión pública que se nutra de la participación de otros sectores de la sociedad, enriqueciendo los procesos de gestión pública e incrementando la capacidad del gobierno para promover niveles de vida más dignos entre la población.

Desde el enfoque de la gobernanza, la evaluación de programas públicos ya no tiene que ser un tema exclusivo del Estado. Se abre espacio para una práctica de evaluación en donde otros actores pueden vincularse, sumando esfuerzos y recursos para alcanzar una política de

evaluación que nos lleve a mejores programas sociales, lo que a su vez traería una mejor satisfacción con la democracia.

En los siguientes apartados se aborda brevemente la satisfacción con la democracia y la eficiencia directiva del gobierno, así como la contextualización para el caso de México.

El desempeño gubernamental y la satisfacción ciudadana con la democracia

El análisis del desempeño gubernamental es de gran relevancia para los Estados latinoamericanos cuyos gobiernos han adoptado recientemente regímenes democráticos, algunos, como México, todavía en proceso de consolidación. El apoyo que la ciudadanía deposita en este sistema político se basa en la percepción que existe sobre su capacidad para solucionar los problemas considerados importantes por la sociedad (Montero, Zmerli y Newton, 2008). Una buena democracia, de acuerdo con Morlino (2004), es un régimen ampliamente legitimado, estable y del cual los ciudadanos están plenamente satisfechos con la calidad de sus resultados.

En este sentido, la insatisfacción política representa un rechazo general de los objetos políticos que no cumplen con los estándares que los ciudadanos establecen para ellos (Montero, Zmerli y Newton, 2008), lo que puede erosionar gradualmente el apoyo que recibe la democracia. Para Torres (1993), es la pérdida de confianza entre los electores y sus partidos, entre la ciudadanía y las políticas estatales (todo lo que queda sin cumplir en una sociedad democrática) lo que produce estados de ingobernabilidad.

Estudios realizados sobre la satisfacción con la democracia abren un debate más amplio sobre la crisis de la democracia representativa ante la variación en los niveles de satisfacción observados (Delgado Sotillos, 2015). En los últimos años, debido a la insatisfacción de grandes sectores de la población en países como Venezuela, Ecuador y Bolivia han surgido y tomado el poder partidos políticos caracterizados por su poco compromiso con la democracia (Vargas Chanes y González Núñez, 2013). Así, entendiendo que la satisfacción con la democracia es determinante para conservar el apoyo de los ciudadanos a este régimen, se han realizado varios estudios para intentar explicar desde diversas perspectivas *la conformación de dicha satisfacción*. En el análisis de factores vinculados a los individuos y a su entorno,

Vargas y Gonzáles (2013) se apoyan de encuestas de percepción sobre variables macro (nivel país) y micro (nivel individual). Montero, Zmerli y Newton (2008), realizan un estudio y determinan que la satisfacción con la democracia se encuentra arraigada en condiciones políticas y en la satisfacción con la vida por parte de la población.

Para Aarts y Thomassen (2008), la satisfacción puede ser explicada a través de la percepción sobre la representación y en menor medida por la rendición de cuentas. En esta línea, Delgado (2015) argumenta que la representatividad de los partidos políticos (políticas implementadas) se explica a través de la receptividad gubernamental (*responsiveness*), un rasgo de la democracia que se circunscribe al proceso representativo y que interconecta a ciudadanos, actores políticos e instituciones a través del proceso electoral (Powell, 2004; citado en Delgado, 2015).

Por otra parte, los resultados de un estudio empírico realizado por Anderson (1998) revelan que los indicadores de desempeño económicos y políticos tienen una mayor capacidad explicativa en los niveles de satisfacción con la democracia que la cultura política. No obstante, para autores como (Wagner, Schneider, y Halla, 2009), la satisfacción con la democracia no solo se centra en los resultados, también es crucial su capacidad para mantener el orden, el respeto de las leyes, el respeto a los derechos humanos y las reglas del juego de la democracia (instituciones de alta calidad).

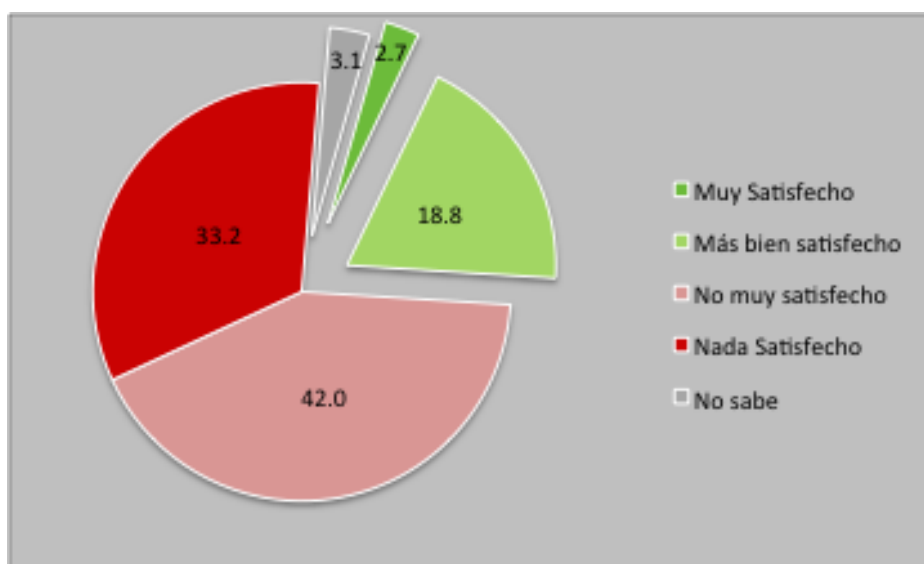
Otro factor que se ha identificado como determinante en la percepción y la evaluación ciudadana es la antigüedad de la democracia, donde las democracias más jóvenes muestran menores niveles de satisfacción (Aarts y Thomassen, 2008; y Anderson, 1998).

Por último, estudios realizados sobre la influencia de la percepción sobre la corrupción en los niveles de satisfacción muestran que, si bien los ciudadanos son críticos respecto al desempeño de las autoridades en términos de corrupción, este criterio no necesariamente se ve traducido en una caída en los apoyos a la democracia (Morales Quiroga, 2009). Sin embargo, los fenómenos de corrupción o escándalos políticos deterioran la calidad del sistema político y erosionan paulatinamente sus niveles de apoyo ciudadano (Kumlin y Esaiasson, 2011; citado en Delgado, 2015).

Sin importar desde que enfoque se estudie la satisfacción con la democracia, el desempeño gubernamental es la variable explicativa de fondo. En este sentido, la evaluación representa una herramienta útil para incentivar un adecuado desempeño, promueve capacidades gubernamentales técnicas orientadas a mejorar la eficiencia y eficacia de la gestión pública que son complementadas con capacidades políticas (como transparencia, rendición de cuentas, e incentivos) que pueden contribuir a mejorar la democracia. En este sentido, una adecuada práctica de evaluación, al ayudar al gobierno a cumplir con sus objetivos, también favorece el logro de mayores niveles de satisfacción ciudadana con la democracia, relación no explorada en la literatura disponible.

En la experiencia mexicana, la evaluación de programas públicos ha cumplido casi una década y no parece impactar en los niveles de satisfacción ciudadana. Según datos del Latinobarómetro para 2013, México era el segundo país con mayor insatisfacción con la democracia en América Latina, después de Honduras. En ese año, el 75 por ciento de la población mexicana se manifestó insatisfecha con los resultados de la democracia (*véase gráfica 3*).

Gráfico 3. Grado de satisfacción con la democracia en México 2013



Fuente: Latinobarómetro.

En países en proceso de consolidación democrática como México, la reivindicación de las demandas sociales y el mejoramiento del desempeño gubernamental para brindar bienes y servicios de calidad, son temas prioritarios. Se requiere de un gobierno con eficiencia

directiva cuyos resultados se vean reflejados en una mayor satisfacción ciudadana, indispensable para mantener el apoyo a este régimen. Un gobierno democrático debe satisfacer las necesidades de la población y para ello podrá acudir a la evaluación como herramienta para lograrlo.

La eficiencia directiva y el desempeño gubernamental

Con la democratización de los regímenes políticos, la legitimidad o la validez institucional han dejado de ser el centro de atención en los estudios y argumentos sobre las limitaciones directivas de los gobiernos. Y en las últimas décadas, el debate se ha centrado en la *capacidad y eficacia* directiva del gobierno, que es abordado desde dos marcos de análisis: en primer lugar, *la gobernabilidad*, desde mediados de los 1970s; y posteriormente, *la gobernanza*, que surge a inicios de la década 1990s (Aguilar Villanueva, 2007).

El término gobernabilidad surge para establecer que la estabilidad o la crisis de la democracia dependen de la probabilidad de que el gobierno *gobierne* a su sociedad. Este enfoque centra su atención en las *capacidades* del gobierno y se sustenta en el supuesto de que “un gobierno capaz, dotado con las capacidades requeridas, es suficiente para dirigir a la sociedad” (Aguilar Villanueva, 2007). Cabe destacar que esta noción presenta dos limitaciones que se confrontan con las condiciones actuales del gobierno: i) su acento en las capacidades, cuando éstas no son más que potencialidades de acción; y ii) una visión donde es el gobierno quien dirige a la sociedad (Aguilar Villanueva, 2007).

Con la atención en las capacidades de los gobiernos, el Banco Mundial, el Banco Interamericano de Desarrollo (BID) y el PNUD han dedicado numerosas publicaciones al estudio del desarrollo de capacidades que permitan mejorar el desempeño del sector público. Principalmente el desarrollo de capacidades nacionales de evaluación, donde se otorga un papel central a la institucionalización de sistemas de Monitoreo y Evaluación debido a que a través de éstos se puede obtener información necesaria para generar procesos de retroalimentación y aprendizaje que permitirán mejorar la gestión pública y los procesos de planeación. En el marco de la GpR, una buena planificación, combinada con un seguimiento y una evaluación eficaz, pueden desempeñar un rol importante para mejorar la eficacia de los programas y proyectos (PNUD, 2009); exige visualizar lo que se espera de la gestión y crear un proceso de planeación así como un instrumento de evaluación para verificar el cumplimiento de las metas y objetivos propuestos (Ospina Bozzi, 2002).

Desde la perspectiva de Aguilar (2007), los resultados de las prácticas de la Nueva Gestión Pública (NGP) que tienen lugar desde finales del siglo XX, llegaron a desarrollar, normalizar y volver socialmente aceptable un estilo de gobernar en modo de gobernanza. En

otras palabras, que las reformas de la NGP, en busca de eficiencia económica rebasaron su perímetro gubernamental-administrativo y detonaron efectos institucionales y políticos que tomaron forma en una nueva relación directiva de los gobiernos con las organizaciones de la sociedad (Aguilar Villanueva, 2007).

La gobernanza⁴ es entendida como el *proceso de gobernar*, a través del cual, el gobierno, las empresas privadas, las organizaciones de la sociedad civil y los ciudadanos interactúan con el propósito de definir, acordar y decidir sus objetivos de interés y valor común, así como acordar y decidir las formas de organización, los recursos requeridos y las actividades a llevar a cabo para lograr los objetivos decididos (Aguilar Villanueva, 2007).

Este concepto tiene dos supuestos: i) la acción del gobierno es necesaria pero insuficiente para la tarea directiva de la sociedad, y por consiguiente ii) lo importante y decisivo es crear un nuevo *proceso de gobernar* más que dotar al gobierno con todas las capacidades requeridas para dirigir a la sociedad. En la literatura hay dos tendencias que se distinguen según su énfasis en la *estructura* o en el *proceso* de la gobernanza (Aguilar Villanueva, 2007).

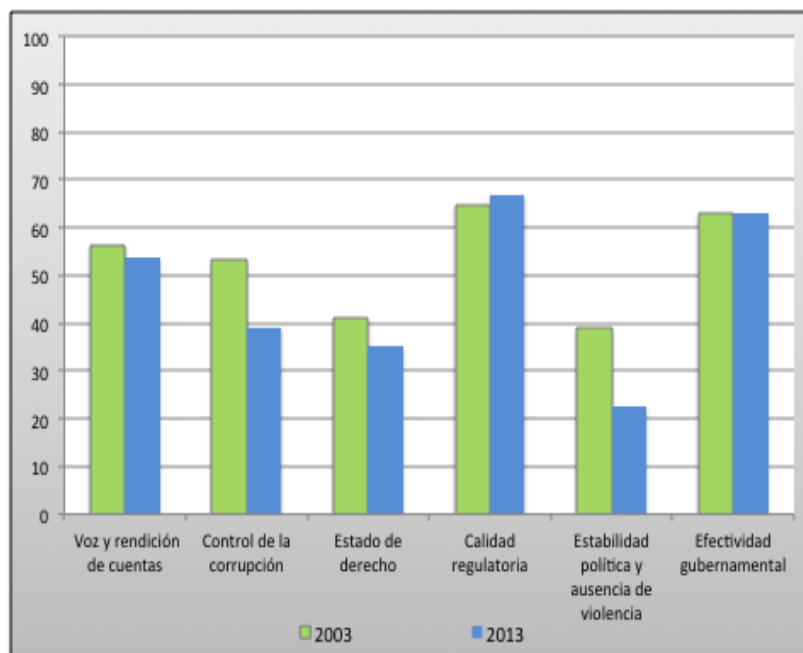
La tendencia que pone el énfasis en la gobernanza como el *proceso de gobernar* reconoce “la importancia de que se respeten valores, principios y leyes del Estado de Derecho y del régimen democrático”; pero brinda mayor atención en la capacidad y eficacia directiva de gobernar, en la producción efectiva de las situaciones sociales deseadas que son exigidas por los valores de la democracia y los derechos de los ciudadanos. Esta corriente orientada a la “la calidad de la democracia” acentúa la dimensión institucional del gobierno democrático (Aguilar Villanueva, 2007). Desde la perspectiva orientada a la *estructura*, Calvo (2010) propone un modelo compuesto por seis variables (legalidad, rendición de cuentas, libertad, igualdad, receptividad y legitimidad) con el propósito de analizar, entender y mejorar la calidad democrática. El autor argumenta que la calidad democrática y la calidad de servicio son dos caras de una misma moneda de la gestión pública, por lo que es necesario el “regreso de la política” que pasaría por un fuerte impulso de renovación y cambio orientado hacia la

⁴ Kaufmann (2005) define la gobernanza como “el conjunto de tradiciones e instituciones mediante las cuales se ejerce la autoridad en un país en pos del bien común. Quedan comprendidos en esta definición el proceso de elección, supervisión y reemplazo de quienes ejercen esa autoridad (la dimensión política), la capacidad del gobierno para gestionar eficazmente los recursos y poner en práctica políticas acertadas (la dimensión económica) y el respeto de los ciudadanos y del Estado por las instituciones del país (la dimensión del respeto institucional)”.

calidad, eficiencia y creación de valor en el funcionamiento de los servicios públicos (Calvo, 2010).

En coincidencia con algunas de las variables que propone Calvo (2010), el Banco Mundial⁵ realiza una estimación anual de seis indicadores de la gobernanza (*governance*) que se ha determinado ayudan a explicar la capacidad de las instituciones nacionales para ejercer la autoridad pública e impulsar el desarrollo económico, social y político duradero (Kaufmann y Kraay, 2008). Al analizar estos indicadores en el transcurso de la última década para el caso mexicano (*véase gráfico 4*), se puede observar que la mitad de éstos registran un decrecimiento significativo: Control de la corrupción, de 53.2 a 39.2 por ciento; Estado de derecho, de 41.1 a 35.1 por ciento; Estabilidad política y ausencia de violencia, de 38.9 a 22.7. De igual forma, pero en menor medida, el indicador de voz y rendición de cuentas muestra un retroceso de casi tres puntos porcentuales. Por otra parte, los dos indicadores que muestran un incremento: calidad regulatoria y efectividad gubernamental presentan una variación apenas perceptible, de 2.3 y 0.3 puntos porcentuales, respectivamente.

Gráfico 4. Indicadores de gobernanza en México, 2003-2013



Fuente: Banco Mundial.

⁵ De acuerdo con un estudio realizado por el Banco Mundial, los aspectos políticos, económicos e institucionales que inciden en la gobernabilidad de un país se pueden integrar por seis componentes: i) Voz y rendición de cuentas; ii) Control de la corrupción; iii) Estado de derecho; iv) Calidad regulatoria; v) Estabilidad política y ausencia de violencia; y vi) Efectividad gubernamental (Kaufmann, Kraay, y Mastruzzi, 2007).

En términos generales, a partir de los planteamientos de Calvo (2010) y los indicadores del Banco Mundial, se podría inferir que a medida que el gobierno mexicano ha retrocedido en la última década en su capacidad para fomentar la gobernanza, también lo ha hecho en su calidad democrática y en su capacidad para impulsar el desarrollo nacional.

El sutil incremento que registra México en el indicador de efectividad gubernamental podría explicarse parcialmente por los esfuerzos de evaluación que ha realizado el gobierno en dicho periodo.

En México, desde la década de los 1970s, el gobierno ha realizado algunos esfuerzos desarticulados para introducir la evaluación de programas públicos en la APF con el propósito de promover las capacidades gubernamentales y hacer frente a los problemas económicos. En años más recientes, desde la instauración de un sistema de gobierno democrático en el país, a inicios del siglo XXI, la evaluación de programas también ha sido orientada a lograr mayores resultados a favor del bienestar de la población. Adicionalmente, esta práctica ha asimilado elementos vinculados a la gobernanza y se ha desarrollado una experiencia de coparticipación Estado y Sociedad; en los procesos de evaluación se ha incorporado la participación de otros actores no gubernamentales (grupos académicos, organismos internacionales, organizaciones de la sociedad civil o consultores privados) a quienes se les ha otorgado la responsabilidad de llevar a cabo las evaluaciones externas de las dependencias, que a su vez son coordinadas por el gobierno.

OBJETIVO GENERAL

Determinar la utilización y el efecto de las evaluaciones de diseño y desempeño realizadas a los programas de Sedesol, entre 2007 y 2014, y si la implementación de las recomendaciones ayudó a los programas a incrementar su capacidad para alcanzar los objetivos.

Objetivos Específicos:

1. Realizar un análisis introductorio que permita un mayor entendimiento sobre cómo se dieron las evaluaciones de diseño y desempeño en Sedesol, y qué tanto fueron utilizadas durante el periodo;
2. Identificar la importancia que tuvieron las características de las evaluaciones y la independencia del proceso evaluativo para la utilización de las recomendaciones;
3. Estimar el efecto que tuvo la utilización instrumental de las evaluaciones en el mejoramiento de los programas de Sedesol a partir de cambios en Reglas de Operación (ROP) y en la estimación del índice de desempeño de GESOC;
4. Examinar a través de dos estudios de caso, los resultados obtenidos en el PDZP y el programa 3X1 a fin de facilitar la interpretación de los hallazgos de apartados anteriores.

PREGUNTAS DE INVESTIGACIÓN

¿La utilización y el efecto de las evaluaciones ayudaron a los programas de la Sedesol a mejorar su diseño y su desempeño, durante el periodo 2007-2014?

Preguntas Específicas:

1. ¿Qué tanto se utilizaron las evaluaciones realizadas durante el periodo?
2. ¿Qué condiciones vinculadas a la práctica evaluativa ayudan a explicar que unas evaluaciones sean más utilizadas que otras?

3. ¿Las recomendaciones implementadas ayudan a explicar los ajustes que experimentaron los programas en sus Reglas de Operación (ROP) y en la estimación de su desempeño?
4. ¿En los estudios de caso es posible observar si la utilización de las evaluaciones deriva en mejoras a la gestión de los programas?

HIPÓTESIS

La utilización y el efecto de las evaluaciones en los programas de Sedesol ha sido limitado, su práctica ha permitido cambios graduales tanto en el diseño como en el desempeño de los programas valorados; sin embargo, éstos no se han traducido en una mejora significativa en su capacidad para solucionar los problemas sociales que les dieron origen.

CAPÍTULO 1

REVISIÓN DE LA LITERATURA SOBRE EL APROVECHAMIENTO DE LAS EVALUACIONES: ANÁLISIS DE LOS FACTORES DETERMINANTES Y LOS EFECTOS

CAPÍTULO 1. REVISIÓN DE LA LITERATURA SOBRE EL APROVECHAMIENTO DE LAS EVALUACIONES: ANÁLISIS DE LOS FACTORES DETERMINANTES Y LOS EFECTOS

La literatura sobre la práctica de evaluación puede dividirse en tres corrientes, según la preocupación central de sus autores: la conveniencia de los *métodos* empleados⁶, la emisión de *juicios de valor*⁷ en torno a los resultados como producto, y el *aprovechamiento* de los resultados. Esta investigación se adscribe a esta última corriente con el propósito de observar en qué medida las evaluaciones ayudan a mejorar la gestión pública.

En este capítulo, presento los principales debates que se han generado sobre el *aprovechamiento* de las evaluaciones señalando la trayectoria que han seguido los investigadores para entender y poder influir favorablemente en esta práctica.

El *aprovechamiento* de la evaluación es un tema de interés desde la década de 1970 y su análisis ha sido tanto responsivo como influyente de las tendencias modernas. Se trata de la corriente de investigación más discutida, cuenta con una importante presencia en la literatura teórica (Christie, 2007; citado en Johnson et al., 2009) y también el mayor número de estudios empíricos (Johnson et al., 2009).

⁶ La corriente que analiza la conveniencia de los *métodos* considera que la evaluación es un ejercicio de investigación social aplicada y que, por ello, debe prestar especial atención a la claridad metodológica y la pertinencia de los modelos empleados para cada caso de estudio (Alkin & Christie, 2004). A finales del siglo XX, al igual que el resto de las Ciencias Sociales, esta corriente supera las discusiones sobre la supuesta superioridad de los métodos cuantitativos respecto a los cualitativos, dejando en claro que los métodos cuantitativos no son siempre los más adecuados para algunos de los propósitos de la investigación o contextos específicos (Reichardt y Cook, 2011).

Actualmente, el debate principal se centra en la conveniencia de los diseños experimentales o cuasi-experimentales, para el caso de investigaciones cuantitativas o mixtas. En esta línea, los trabajos de Donald Campbell se consideran de gran relevancia, en particular, el clásico de Campbell y Stanley (1966) donde se definen las condiciones básicas que deben cumplir estos diseños para contar con rigor académico (citado en Alkin y Christie, 2004). De la contribución de Campbell y Stanley se han derivado numerosas propuestas de nuevos diseños.

⁷ La corriente que estudia los *juicios de valor* implícitos en el ejercicio de evaluación hace énfasis en la distinción entre los evaluadores y otros investigadores en ciencias sociales, argumentando que los primeros reflejan su sistema de creencias y valores en los resultados que presentan al determinar qué es bueno y qué es malo. De acuerdo con Merino (2011), si las evaluaciones son una valoración, ésta solamente puede lograrse en función de los valores implícitos que cada intervención gubernamental dice defender (ética pública), así como de los valores que deben seguir los servidores públicos responsables de llevarla a cabo (ética administrativa). Desde esta perspectiva, la evaluación entendida como una cuestión ética representa un ejercicio fundamental para revisar la coherencia entre las políticas (con los valores que invocan) y los medios que emplean (Merino, 2011, 178).

El principal exponente de esta corriente es Michael Scriven, cuyos primeros trabajos (1967 y 1972) definen, en gran medida, los fundamentos conceptuales y procedimientos de la disciplina (Alkin y Christie, 2004).

Esta corriente surge ante la necesidad de verificar si los resultados de las evaluaciones son empleados y si efectivamente ayudan a cumplir los objetivos previstos. El argumento de los estudios que se realizan sobre el *aprovechamiento* de los resultados, se centra en que la legitimidad de las evaluaciones radica en su beneficio práctico: su habilidad para mejorar políticas, programas y proyectos (Patton, 1997; citado en Ledermann, 2012); y por tanto, que si una evaluación no se emplea, sin importar su calidad, puede considerarse un fracaso (Grob, 2003 y Patton, 1997; citado en Herbert, 2014).

El nombre con que comúnmente se conoce esta corriente en la literatura es “uso de la evaluación” (*use of evaluation*). Sin embargo, para fines del presente trabajo, sustituyo la palabra *uso* por *aprovechamiento* para evitar confundir la corriente con los marcos de análisis que se desprenden de ella, ya que son nombrados con una terminología muy similar: el *uso*, la *utilización* y la *influencia*.

Este capítulo se compone de cuatro apartados. En el primero, menciono brevemente las principales aportaciones de los tres marcos de análisis: el *uso*, que analiza exclusivamente la aplicación de las recomendaciones para la toma de decisiones; la *utilización*, que explora además el efecto de las recomendaciones aplicadas por los actores vinculados al proceso; y la *influencia*, que observa tanto los efectos directos como los indirectos de los resultados, en actores vinculados y no vinculados al proceso evaluativo. Para la presente investigación empleo los dos últimos: la *utilización* (efecto inmediato) y la *influencia* (efecto no inmediato).

En la segundo apartado, desde el marco de análisis de la *utilización*, retomo los modelos desarrollados por Cousins y Leithwood (1986) y Balthasar (2009) para identificar los factores que determinan la aplicación de las recomendaciones y, posteriormente, los clasifico en tres planos de análisis, según su cercanía al proceso evaluativo: el *endógeno*, referente a las características de la evaluación en su implementación; el *exógeno inmediato*, que aborda las condiciones bajo las cuales se toman las decisiones o se configura la intervención gubernamental; y *los arreglos institucionales*, responsables de la autonomía del proceso evaluativo. También abordé la evidencia empírica sobre la relevancia de los factores *endógenos* e *institucionales* ya que la empleo en esta investigación para establecer hipótesis sobre el *aprovechamiento* de las evaluaciones en el caso mexicano.

En el tercero, desde el marco de análisis de la *influencia*, retomo los modelos de Kirkhart (2000), Alkin y Taut (2003), y Henry y Mark (2003 y 2004) para identificar los ámbitos de la gestión pública en los cuáles la evaluación puede impactar, directa o indirectamente, en el diseño y en el desempeño de los programas sociales.

Por último, en el cuarto, menciono algunos de los estudios empíricos que se han realizado en México a partir de los planteamientos teóricos desarrollados en torno al *aprovechamiento* de la evaluación. Estos estudios se han centrado principalmente en la *utilización* de los resultados y más recientemente en su *influencia* directa en los programas y en la toma de decisiones.

1.1. EL DESARROLLO DE LOS DEBATES SOBRE EL APROVECHAMIENTO DE LA EVALUACIÓN

En este apartado, señalo brevemente el desarrollo de la corriente académica que analiza las evaluaciones a partir del *aprovechamiento* de los resultados. Para tal efecto, los principales debates e investigaciones que se desprenden de esta corriente son clasificados en tres marcos de análisis identificados por Herbert (2014) como: el *uso* (la aplicación de las recomendaciones), la *utilización* (los efectos directos de las recomendaciones aplicadas), y la *influencia* (los efectos directos e indirectos de los resultados).

Cabe mencionar que el auge académico de los tres marcos de análisis identificados (*uso*, *utilización* e *influencia*), se da en diferentes periodos, ya que representan etapas de una misma corriente, que se desarrolla desde la década de 1970 a la actualidad. Son tres enfoques que buscan explicar un mismo fenómeno: *el aprovechamiento* de la evaluación, y emplean variaciones en la terminología con el propósito de señalar que hay diferencias en la manera en que abordan su objeto de estudio. El *uso* de la evaluación, 1971-1985, se restringe a observar si las recomendaciones de la evaluación son consideradas (o no) para la toma de decisiones. La *utilización*, 1986 - 2000, sustituye el término *uso* por *utilización* al mismo tiempo que amplía el campo de estudio; entonces, introduce al análisis el impacto de las recomendaciones en programas, organizaciones o actores vinculados al proceso. Por último, la *influencia*, a partir el año 2000, analiza el *aprovechamiento* en términos todavía más amplios reconociendo que la práctica evaluativa puede tener efectos más allá de los ámbitos controlados o pertinentes para los evaluadores y los evaluados.

A continuación, retomo el trabajo de Herbert (2014) y lo complemento con el de Alkin y Christie (2004)⁸ para señalar brevemente las principales aportaciones de cada marco de

⁸ Alkin y Christie (2004), en vez de realizar una revisión histórica como Herbert (2014), analizan las aportaciones más relevantes sobre el *aprovechamiento* de la evaluación a través de los exponentes más destacados en este campo de estudio, estableciendo tres categorías: i) el uso en la toma de decisiones (en la que destacan autores como Daniel Stufflebeam, Malcolm Provus, Joseph Wholey, Marvin Alkin); ii) la utilización de los resultados y su impacto (destacando Michael Patton, Marvin Alkin, Bradley Cousins, Hallie Preskill, Jean King, John Owen); y iii) la evaluación como medio de empoderamiento (David Fetterman).

Cabe mencionar que la clasificación que establecen estos autores coincide en gran medida con los marcos de análisis identificados por Herbert (2014). Las primeras dos vertientes establecidas por Alkin y Christie (2004) y Herbert (2014) guardan gran semejanza; sin embargo, la última vertiente referida a la evaluación como medio de empoderamiento representa solo una de la numerosas propuestas que tienen lugar en el último periodo identificado por Herbert (2014), donde se reconoce en términos amplios los efectos indirectos que las evaluaciones pueden generar.

análisis. En cada caso, menciono el enfoque a través del cual se analiza el *aprovechamiento* de las evaluaciones; y también, el tipo de estudios empíricos que se han realizado para entender la ocurrencia (o no) de este fenómeno.

EL USO (1971 – 1985)

Este fue el primer marco de análisis que se desarrolló para entender el *aprovechamiento* de las evaluaciones, sentando las bases de la corriente académica. Durante el periodo, las investigaciones realizadas conceptualizan el *uso* de la evaluación en términos de su aplicación, o no, en la toma de decisiones (Herbert, 2014, p. 389). Según el efecto esperado, se distinguieron diferentes tipos de usos: instrumental, conceptual y simbólico (Sager & Ledermann, 2008).

En el *uso instrumental*, las recomendaciones sirven para informar la toma de decisiones y para generar cambios en la intervención pública valorada; este es el único tipo de uso que motiva y justifica en términos formales las iniciativas de evaluación. En el *uso conceptual*, los resultados de la evaluación se aplican para un mejor entendimiento o un cambio en la concepción del objetivo de la evaluación. Y por último, en el *uso simbólico*, los resultados de la evaluación son empleados para justificar o legitimar una posición preexistente, sin efectivamente cambiarla (Johnson et al., 2009; Sager & Ledermann, 2008; Herbert, 2014). Esta es una clasificación que hoy en día se mantiene vigente.

El *uso*, en general, fija como objetivo emplear los hallazgos de los evaluadores para tomar decisiones más sabias, ayudando a los responsables a decidir si un programa debe terminar, ser ampliado, modificar sus actividades o cambiar la formación su personal (Weiss, 1998). Una debilidad importante de este enfoque fue asumir que al producirse información relevante, basada en resultados a través de las evaluaciones, ésta impactaría de manera casi automática en los programas y en las decisiones de los responsables sobre lo que se debería hacer a continuación (C. Weiss, 1998).

En este sentido, la evidencia ha dejado claro que las recomendaciones de la evaluación no se aplican invariablemente para mejorar las intervenciones públicas. Y a pesar de que la

razón principal por la que se evalúa es proporcionar información útil para la toma de decisiones, los responsables de programas no suelen aceptar fácilmente los resultados y, por tanto, no siempre los emplean para mejorar sus intervenciones (House, 1992).⁹

Podría pensarse que existían expectativas poco realistas sobre la capacidad de la evaluación para solucionar los problemas de la gestión gubernamental y que los partidarios de esta práctica se vieron obligados a reconocer sus limitaciones. Fue necesario tomar en cuenta que la evaluación puede entregar recomendaciones e informar una decisión pero no necesariamente llegar a cambiarla; y además, que un actor puede tener comúnmente buenas razones para ignorar los hallazgos obtenidos en el proceso evaluativo (Birkland, Murphy-Graham & Weiss; citado en Herbert, 2014, p. 390). Esto debido, en gran medida, a que las evaluaciones son sólo una de las numerosas fuentes de información que deben considerar los tomadores de decisiones en sus ámbitos de acción (Weiss, 1987; citado en Herbert, 2014, p. 390).

Por otra parte, además de los problemas que puede enfrentar la práctica evaluativa por sí misma. A finales de la década de los setenta, Carol Weiss fue de los primeros investigadores en exponer la posibilidad de que, en algunos casos, las evaluaciones no sean aprovechadas de la manera prevista (observándose el no uso o mal uso de los resultados), por responder a motivaciones no legítimas que la autora denomina como “encubiertas” y que son consideradas nocivas para esta disciplina.

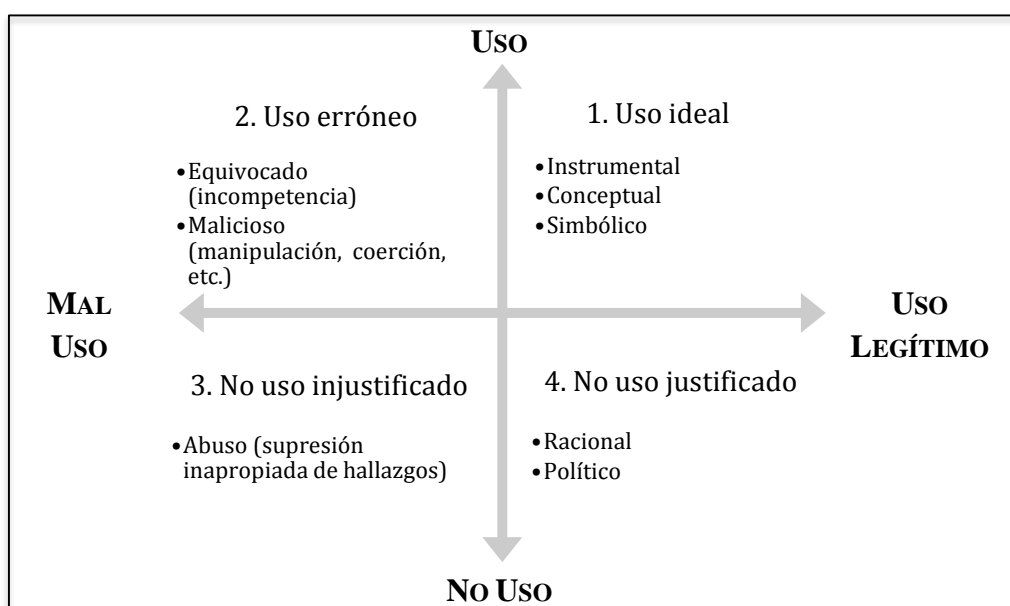
Al respecto, Weiss (1972), señala que los motivos por los cuales se decide llevar a cabo una evaluación son diversos, desde los racionales hasta los políticos y que, aunque idealmente un administrador debería buscar respuestas sobre el futuro del programa a evaluar, en algunas ocasiones esta práctica se adopta por razones menos legítimas como: aplazar decisiones,

⁹ Reconociendo el hecho de que los hallazgos de la evaluación no siempre logran impactar las intervenciones públicas que valoran, se han dedicado numerosos trabajos a entender el *no uso* y el *sub-uso* de la evaluación. Al respecto Carol Weiss (1972), señala que un limitado uso de la evaluación puede estar vinculado al rol que desempeña el evaluador (en veces más interesado en hacer un trabajo que sea de interés para su profesión que dar respuesta a las cuestiones requeridas por los administradores) o a la resistencia al cambio por parte de los organizadores (por razones de factibilidad, aceptabilidad o ideología). Asimismo, Teresa Bracho (2011), considera que el limitado uso de la evaluación se explica, en gran medida, por el desacuerdo entre la agenda pública y la de la investigación en relación con qué investigar, cómo usar los resultados y para qué sirve el conocimiento de las políticas públicas; de tal forma que el uso de la evidencia científica suele complicarse cuando: hay bajo valor de la evidencia, no se responde a las necesidades, los estudios cualitativos son poco útiles para los políticos, se generan dilemas en la interpretación sobre las consecuencias, hay un desfase entre los tiempos políticos y los académicos, o la realidad es demasiado compleja.

esquivar responsabilidades, fortalecer relaciones públicas, mejorar su prestigio, o el mero cumplimiento de los requisitos de la subvención, entre otras.

A partir de estos hallazgos, Cousins (2004) elabora un esquema donde señala cuatro posibles escenarios en torno al *uso* (uso ideal, mal uso, no uso justificado y no uso injustificado), los cuales reflejan las actitudes que pueden adoptar los usuarios sobre la implementación de los resultados, ya sea que éstos sigan motivaciones legítimas o no (*véase ilustración 1.1*).

Ilustración 1. 1 Posibles escenarios para el uso de los hallazgos de la evaluación



Fuente: Elaborado por Cousins, 2004.

Los estudios empíricos derivados de este enfoque se limitaban a analizar el *uso* de la evaluación tomando en cuenta exclusivamente si las recomendaciones eran atendidas o no (Herbert, 2014, p. 389). Al respecto, Leviton y Boruch (1983), afirman que resultaba más fácil analizar el uso de las recomendaciones que otra cosa, y que por ello los estudios a menudo solo contaban el número de recomendaciones que habían sido aplicadas (en Weiss, 1998, 24).

Esta limitación generó importantes críticas por los débiles estándares de evidencia, particularmente por la falta de medios de verificación para determinar si se generaban cambios en el comportamiento a partir de las recomendaciones atendidas (Leviton, 2003; citado en Herbert, 2014).

Para finales de este periodo se abandona el enfoque simplista con que se inicia el estudio del *aprovechamiento* en la década de 1970, y hay un acuerdo generalizado de que el *uso* no es un concepto unitario sino un fenómeno multidimensional (Leviton y Hughes, 1981, citado en Shulha y Cousins, 1997, p. 196). Se reconoce la necesidad de integrar en su análisis elementos ajenos al proceso evaluativo que pueden afectar de manera importante esta práctica (p. ej. La existencia de otra fuente de información relevante para quienes toman decisiones, el entorno en que se ejecuta el programa valorado, entre otros). En este sentido, de acuerdo con Shulha y Cousins (1997; citado en Herbert, 2014), cuando concluye este periodo, el desarrollo más relevante que se genera para el entendimiento del *uso* es la creciente importancia que se le atribuye al contexto; específicamente cuando los investigadores de la evaluación empiezan a preocuparse por cómo los resultados interactúan con otros elementos en los procesos de toma de decisiones y cómo los evaluadores deben enganchar en éste.

Además, se promueve la evaluación como un diálogo entre los evaluadores y los participantes del programa con el propósito de fomentar una mayor interacción y de compartir la responsabilidad de generar, transmitir y emplear la información de la evaluación. Esto renovó el entusiasmo por el potencial de la evaluación y por una nueva década de investigación conceptual y empírica (Shulha y Cousins, 1997).

LA UTILIZACIÓN (1986 – 2000)

A mediados de la década de 1980, las limitaciones en el estudio del *uso* de las evaluaciones (considerado vago e inconsistente en la literatura) propician el surgimiento de un nuevo marco de análisis: la *utilización*. Entonces, las investigaciones sobre el *aprovechamiento* de las evaluaciones se centran no sólo en la aplicación de los resultados sino también en su impacto en programas, organizaciones o actores vinculados al proceso. Desde este enfoque, la *utilización* es entendida como la aplicación directa de los hallazgos de manera que hay efectos observables, inmediatos, concretos, así como cambios paulatinos en actitudes o puntos de vista dentro de una organización (Husseiny y Earl, 1996).

El surgimiento de este marco de análisis generó un debate importante sobre lo que constituye la *utilización* y cómo se puede medir, enfrentándose diferentes posturas respecto al

alcance que se le debe atribuir al término. De acuerdo con House (1992), uno de los eventos más importantes en este periodo fue el intenso debate académico entre Carol Weiss y Michael Patton, quienes difieren sobre en qué medida los resultados de las evaluaciones son aplicados o no, para generar cambios en el objeto de análisis (desde una perspectiva instrumental)¹⁰ y hasta qué punto los evaluadores son responsables de ello.

Por una parte, Carol Weiss argumentaba que esperar la revisión de los programas como resultado directo de las evaluaciones es irrealista. Para la autora, el proceso de *utilización* es un proceso de interpretación y esclarecimiento a través del cual los resultados se van incorporando gradualmente a las estructuras generales de referencia de los usuarios. Teniendo en consideración la toma de decisiones organizacional, Weiss anima a los evaluadores a aspirar a una utilización instrumental pero no emplear este criterio para medir el éxito o la calidad de una evaluación (House, 1992; y Herbert, 2014).

Michael Patton, por otra parte, insistía en la validez del uso instrumental afirmando que los resultados serán aplicados si son presentados de forma apropiada. Patton consideraba que los evaluadores deben vincularse activamente para fomentar la *utilización*, esto implica generar información y procesos que atienden las necesidades de los responsables de la evaluación (Herbert, 2014). Desde esta perspectiva, el trabajo de un evaluador no se limita a producir hallazgos sino a generar un efecto (Patton, 2008), promoviendo la creación de incentivos para que las recomendaciones se implementen (Lawrenz, Gullickson & Toal, 2007; citado en Herbert, 2014, p. 389).¹¹

A inicios de este periodo existían numerosos estudios empíricos que buscaban explicar la utilización de las evaluaciones, a través de variables que comúnmente se agrupan en categorías tales como: a) la pertinencia, b) la credibilidad, c) la participación de los usuarios,

¹⁰ Stame (citado en Msila & Setlhako, 2013) señala que las evaluaciones basadas en la teoría de cambio han ayudado a abrir la "caja negra" de los programas. Este autor define la caja negra como el espacio entre los insumos reales y los resultados previstos del programa.

¹¹ En este periodo, Patton (1998), también incorpora al debate un nuevo momento para el aprovechamiento de la evaluación, la utilización durante el proceso. Se trata de una propuesta que suma la experiencia de participantes involucrados en la evaluación, defendiendo la idea de que cambios pueden ocurrir durante el proceso evaluativo y convirtiendo el aprendizaje de los participantes también en un objetivo de la evaluación, además de informar a los tomadores de decisiones (citado en Herbert, 2014). Esta última propuesta de Patton ha influido tanto en la teoría como en la práctica de evaluación, de tal forma que la noción "utilización durante el proceso" se puede ver reflejada en modelos recientes sobre los efectos de la evaluación. Motivando en parte, los acercamientos modernos para procesos de evaluación participativos o colaborativos (Patton, 1997, Herbert, 2014; House, 1992).

d) la eficacia de la comunicación, e) el potencial para el procesamiento de la información, f) la necesidad de los clientes para obtener información, g) el grado de cambio previsto, h) el valor percibido de la evaluación como herramienta de gestión, i) la calidad de la implementación de la evaluación, y j) las características contextuales del entorno de las decisiones o política (Shulha y Cousins, 1997).

Un trabajo de gran relevancia fue el de Cousins y Leithwood (1986), donde se realiza un meta-análisis de estudios empíricos previos para explorar la importancia relativa de cada factor en su capacidad para predecir la *utilización*. Los autores emplean un índice de “prevalencia de relaciones” y concluyen que la calidad de la evaluación es la característica más importante; seguida de las características de las decisiones, la receptividad a la evaluación, los hallazgos y la relevancia (citado en Johnson et al., 2009).

Este meta-análisis ha sido ampliamente citado y cuenta hoy en día con gran vigencia. Johnson et al. (2009), replican este estudio para el periodo 1986-2005, obteniendo como resultado que el compromiso, la interacción y la comunicación, entre los usuarios de la evaluación y el evaluador, son críticos para la *utilización*.

Shulha y Cousins (1997) afirman que éste es uno de los pocos trabajos que reconocen que no basta con describir los diferentes tipos de *uso* y catalogar los factores determinantes, y que desafían a la comunidad de investigación en su esfuerzo por asistir la construcción de teorías. Además, señalan que a pesar de su relevancia, no ha habido más esfuerzos significativos en esta dirección.

Otro trabajo importante, elaborado en años más recientes, es el de Balthasar (2009) quien introduce al análisis seis factores institucionales (regulación sobre el actor detonante, la unidad responsable de la evaluación, el origen del presupuesto, la contratación, el tipo de evaluador, y la comunicación de resultados) con la intención de explorar la independencia del proceso evaluativo y su relevancia para la *utilización*. El autor concluye en su trabajo que la distancia entre el evaluador y el evaluado es importante pero no siempre determinante.

Una propuesta más reciente que mezcla algunos de los factores mencionados es la de Cherly Gray (2016), quien, basándose en su experiencia en organizaciones internacionales, sugiere un marco conceptual mediante el cual explica qué se necesita para que las

evaluaciones sean utilizadas. Dos aspectos novedosos de esta propuesta son: a) distingue entre los factores vinculados de la oferta (evaluadores) y los de la demanda (usuarios) y b) la calidad se analiza con las características del personal evaluador, sus capacidades y los incentivos a los que están sujetos.

Cuadro 1. 1 Marco conceptual propuesto por Cherly Gray sobre los factores relevantes para la utilización de las evaluaciones

OFERTA (EVALUADORES)	DEMANDA (USUARIOS)
Entorno institucional Recursos Suficientes Grado de Independencia Acceso a Información	Capacidad Habilidad y recursos de una organización para aprender y cambiar.
Calidad Buen Personal Capacitación (metodología, comunicación, juicio, conocimiento sobre el sector) Incentivos para los evaluadores (enfrentar conflicto de intereses en contratación)	
Tono y accesibilidad (recomendaciones honestas pero constructivas, vincular a los evaluados pero mantener distancia)	Entorno de gobierno Motivación de aquellos con poder sobre una organización para aprender y cambiar
Coordinación (pertinencia de los tiempos)	
Esfuerzos de diseminación	

Fuente: Presentación de Cherly Gray en el Seminario Internacional de estudios empíricos sobre la utilización de la evaluación.

Más adelante, en la segundo apartado de este capítulo, retomo los trabajos de Cousins y Leithwood (1986); y Balthasar (2009), para abordar a mayor profundidad el análisis de la *utilización*, a partir de factores explicativos (*véase apartado 1.2*).

LA INFLUENCIA (2000 – PRESENTE)

Desde inicios del siglo XXI aproximadamente, se ha desarrollado un tercer marco de análisis: la *influencia*. En términos amplios, se trata de un acercamiento nuevo y todavía en desarrollo que busca explicar el *aprovechamiento* de la evaluación y sus efectos.

Este nuevo enfoque surge por la inconformidad de algunos académicos con las limitaciones conceptuales del *uso* y la *utilización*. Al respecto, Henry y Mark (2003) y Kirkhart (2000) argumentan que los otros marcos de análisis subestiman el impacto de la

evaluación al enfatizar una aplicación basada en resultados, y al establecer que los efectos siempre son intencionados, inmediatos y directos.

Para facilitar la comprensión de la *influencia* se ha desarrollado un cuerpo académico y de investigación, en un lapso relativamente corto de tiempo, rescatando los trabajos de los últimos 30 años sobre el *aprovechamiento* de las evaluaciones. De acuerdo con Herbert (2014), la comunidad académica de esta disciplina ha comenzado a debatir la *influencia* como un posible sucesor de la *utilización*, ya que ofrece un marco de referencia en el cual se consideran los efectos previstos e imprevistos que la evaluación puede tener, lo que es particularmente importante considerando las metas más indirectas de la evaluación moderna (el aprendizaje organizacional y el empoderamiento) y la responsabilidad de los evaluadores de contribuir a mejorar las condiciones sociales a través de su trabajo.

En la literatura, la aplicación del término *influencia* ha sido propuesta tanto como una extensión (Alkin & Taut, 2003; Kirkhart, 2000), como un remplazo de la *utilización* (Mark & Henry, 2003-2004). En ambos casos, de acuerdo con Kirkhart (2000), la principal diferencia consiste en que la *influencia* representa un acercamiento para estudiar los efectos de la evaluación en los términos más amplios posibles (citado en Herbert, 2014).

El análisis de la *influencia* se basa en la idea de que, en ocasiones, a pesar de que los resultados no son empleados en la manera en que los evaluadores esperan, éstos pueden tener un impacto considerable. En esta dirección, investigaciones han mostrado que sin importar la renuencia de los tomadores de decisiones para aplicar las recomendaciones emitidas, los resultados frecuentemente impactan en decisiones sobre la gestión o las prácticas de los programas (Alkin et al. 1979; Becker, Kirkhnart & Doss, 1982), al cambiar la forma en que los problemas son entendidos (Caplan et al., 1975).

Algunos académicos renuentes a la adopción de este término como Alkin y Taut (2003), Hofstetter y Alkin (2003), y Patton (2008), sugieren que el análisis de la influencia no es de ayuda debido a que considera eventos y factores que están más allá del conocimiento o el control del evaluador (citado en Herbert, 2014). Además, Hofstetter y Alkin (2003) y Patton (2008) enfatizan que el propósito de la evaluación es el mejoramiento de los programas y el mejoramiento social. Por tanto, el estudio de los alcances indirectos de esta práctica no es considerado relevante. No obstante, para Herbert (2014), si bien algunos elementos de la

influencia no pueden ser controlados por el evaluador, es igualmente valioso entenderlos y desarrollar un cuerpo de literatura que los reconozca.

El análisis de la *influencia* es ampliamente discutido. Sin embargo, en la literatura existen todavía muy pocas investigaciones donde se explora directamente este enfoque, observándose una lenta acumulación de evidencia empírica (Herbert, 2014).

En la presente investigación, el análisis de la *influencia* no reemplaza el de la *utilización*, sino que lo complementa para observar el efecto de la evaluación una vez que las recomendaciones han sido aplicadas.

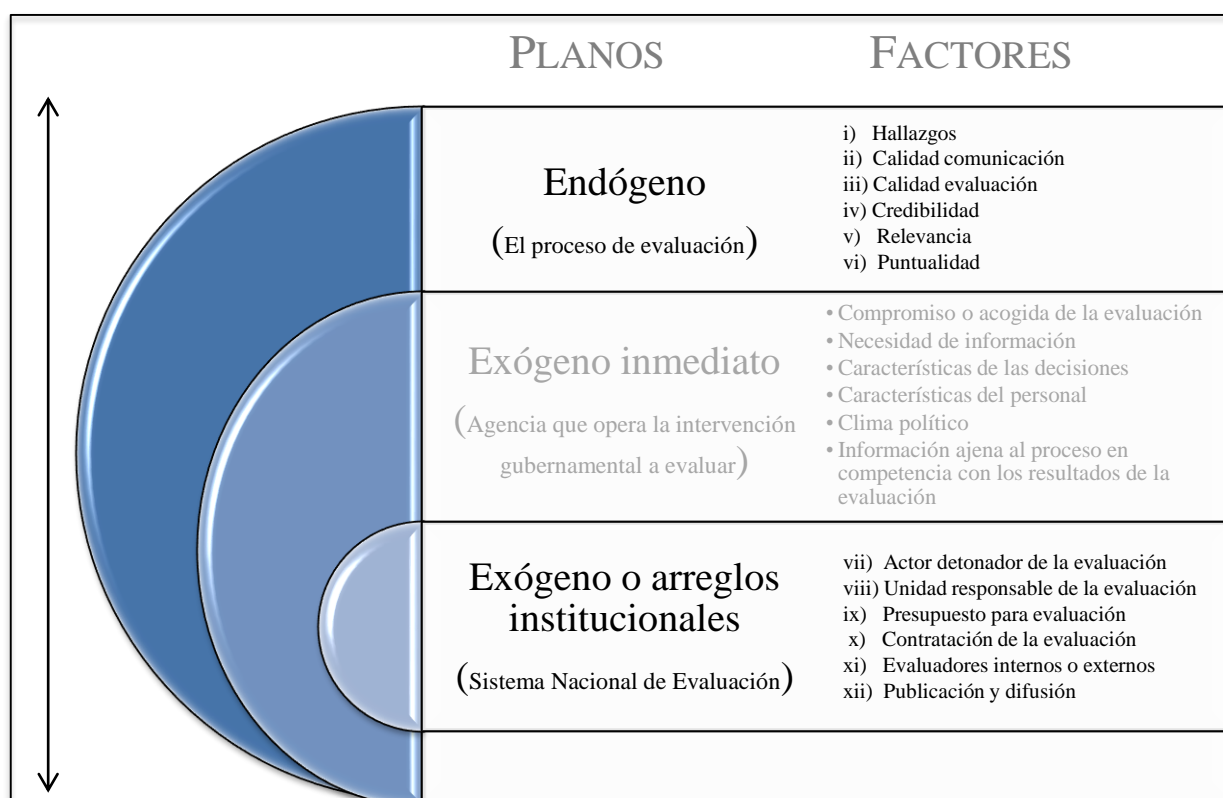
En el tercer apartado de este capítulo, con base en el modelo propuesto por Mark y Henry (2004), identifiqué diversos mecanismos generadores de cambios que se pueden desprender de la práctica evaluativa e impactar en la gestión pública (*véase apartado 1.3*).

1.2. ¿QUÉ FACTORES DETERMINAN SI LAS EVALUACIONES SE UTILIZAN?

En este apartado, con el propósito de identificar variables explicativas que me permitan establecer hipótesis sobre el *aprovechamiento* de las evaluaciones, retomo las principales propuestas teóricas sobre la *utilización*, así como los hallazgos de algunos estudios empíricos.

Siguiendo la línea de investigaciones que buscan entender y controlar las variables que afectan la *utilización*, retomé los factores predictivos identificados en los modelos de Cousins y Leithwood (1986), y Balthasar (2009). Posteriormente, los clasifiqué en tres planos para analizar la relación que guardan con la *utilización*, a partir de su cercanía al proceso evaluativo: i) el endógeno, ii) el exógeno inmediato, y iii) los arreglos institucionales (véase *Ilustración 1.2*).

Ilustración 1. 2 Factores determinantes en la utilización de las evaluaciones de intervenciones gubernamentales, según su cercanía al proceso evaluativo



Fuente: Elaboración propia a partir de Cousins y Leithwood (1986), y Balthasar (2009).

El **plano endógeno o arreglos institucionales** se compone de seis factores que reflejan las *características de la evaluación en su implementación*; el **exógeno inmediato** de seis que revelan las *condiciones bajo las cuales se toman las decisiones o se configura la intervención gubernamental*; y por último, los **arreglos institucionales** de factores que determinan la *independencia del proceso evaluativo*.

Los factores endógenos y los exógenos inmediatos pueden ayudar a explicar la *utilización* de los resultados a través de las particularidades de cada proceso evaluativo. Cousins y Leithwood (1986), analizan en su trabajo la relevancia de estos dos tipos de factores. Para ello, los autores revisaron 65 investigaciones empíricas elaboradas entre 1971 y 1985, y efectuaron un meta-análisis que les permitió reconocer variables explicativas de diferente naturaleza, establecer una clasificación y estimar su relevancia. Su trabajo, a diferencia de la mayoría de aquellos realizados hasta la fecha, no pretendía solo identificar variables capaces de predecir la utilización de la evaluación, sino también estimar la relevancia relativa de cada una de esas variables consideradas.

Cousins y Letihwood concluyen que la relevancia de los factores observados varía, siendo más evidente cuando: a) las evaluaciones son apropiadas en enfoque, sofisticación metodológica e intensidad; b) las expectativas de los usuarios fueron consistentes con la de los hallazgos; c) los usuarios se involucran en el proceso de evaluación y tienen el compromiso con los beneficiarios de la evaluación; d) los usuarios consideran que la información reportada en la evaluación es relevante para sus problemas; e) una cantidad mínima de información de otras fuentes está en conflicto con los resultados de la evaluación.

Debido al éxito que tuvo dicha investigación entre los estudiosos de esta corriente académica, Johnson (2009) replicó el meta-análisis para el periodo 1986 - 2005. El autor emplea el mismo marco de análisis que el primer trabajo y revisa 46 estudios empíricos. El principal aporte de esta nueva investigación fue incorporar al estudio de la *utilización* nuevos elementos: la capacidad del evaluador para la implementación de la evaluación y el compromiso de los tomadores de decisiones con el proceso evaluativo. En esta tesis, de los dos elementos que incorpora Johnson, únicamente considero para el análisis la capacidad del evaluador.

Por otra parte, los factores del tercer plano de análisis los retomé del modelo desarrollado por Balthasar (2009), quien motivado por el creciente interés académico en la institucionalización del proceso evaluativo, explora la relevancia de los arreglos institucionales en la *utilización* a partir de 300 evaluaciones realizadas por la Administración Federal de Suiza entre 1999 y 2002.

Balthasar (2009) define los arreglos institucionales como “la organización formal de los procesos y las competencias, así como las normas de procedimientos que son aplicables independientemente de los proyectos individuales de evaluación”; y afirma que estos reflejan la práctica evaluativa de una institución al establecer la distancia entre evaluadores y evaluados. En este sentido, que a medida que aumenta la distancia entre los actores, mayor es la autonomía del proceso evaluativo.

Cabe mencionar que los factores exógenos de la práctica evaluativa, referentes a las condiciones bajo las cuales se toman las decisiones o se configura la intervención gubernamental (segundo plano de análisis) no se consideran en este trabajo ya que se realiza un estudio retrospectivo y sería muy complicado obtener información confiable. Esto debido, por un lado, a que en la mayoría de los casos (dada la naturaleza política de los mismos) dichos factores no dependen ni quedan plasmados en fuentes oficiales; y por otro lado, a razón de la elevada rotación de personal que experimenta la administración pública mexicana cada cambio de gobierno. El hecho de que ya no ocupen sus puestos las personas que estuvieron vinculadas a los procesos evaluativos en años anteriores, limita el acceso a testimonios fidedignos sobre lo acontecido.

A continuación, presento los principales resultados de los meta-análisis que han explorado la relevancia de los factores (endógenos e institucionales) en la *utilización* de las evaluaciones.

1.2.1. FACTORES ENDÓGENOS: LAS CARACTERÍSTICAS DE LA EVALUACIÓN

En esta sección retomo las meta-evaluaciones realizadas por Cousins y Leithwood (1986); y Johnson (2009).

- i) *Hallazgos*. Este factor ha sido definido en función de la naturaleza de los resultados (positivos o negativos), el grado de congruencia con las expectativas de la audiencia o su valor para la toma de decisiones.

La mayoría de los estudios concluyen que cuando los resultados de la evaluación son congruentes con las expectativas de los tomadores de decisiones, la aceptación y la utilización aumenta; y que cuando los hallazgos son incongruentes tienden a ser ignorados y otra información se utiliza en su lugar.

También se encontró que cuando los hallazgos son prácticos, concluyentes e identifican cursos alternativos de acción tienden a ser más utilizados; principalmente para fines legislativos, desarrollo organizacional y la toma de decisiones federales o locales (Alkin et al. 1974; Carter, 1971; David, 1978; Florio et al. 1979; Leviton y Boruch, 1983; Rich, 1979; Sproull y Zubrow, 1981).

En el análisis de Johnson (2009), seis estudios empíricos exploraron la relevancia de los hallazgos en la *utilización* y obtuvieron conclusiones mixtas. Solo en dos estudios fueron importantes, aunque menos que la comunicación, la puntualidad y la calidad de la evaluación.

- ii) *Calidad en la comunicación*. Este factor es definido a partir de la claridad con que se reportan los resultados de la evaluación a la audiencia: en términos de estilo, la argumentación de los resultados por parte del evaluador, la frecuencia de los informes de resultados o la amplitud de su difusión. En algunos casos, también se ha considerado el proceso de comunicación entre los evaluadores y los clientes.

Una tercera parte de los estudios empíricos revisados por Cousins y Leithwood (1986) se interesaron por la relación entre la *utilización* y la comunicación, pero la evidencia no fue concluyente. Por un lado, algunas investigaciones reportaron que los esfuerzos por

realizar presentaciones orales de los resultados, a la par de informes escritos, contribuyen a resultados de mayor impacto y legibilidad, logrando mayor conciencia y apreciación sobre los resultados (Breckner et al. 1982; Bigelow y Ciarlo, 1976; Rossman et al. 1979; Tash y Stähler, 1982; Van de Vail y Bolas, 1982).

No obstante, otro grupo de investigaciones reportó que el estilo de la comunicación había demostrado tener poca o nula influencia sobre las decisiones del programa, el conocimiento de los resultados y otras medidas de *utilización* (Alkin et al., 1974; Bigelow y Ciarlo; Dickey, 1980; Florio et al., 1979).

En el análisis de Johnson (2009), once estudios abordaron la calidad en la comunicación y la mayoría encontraron que este factor tiene una relación con la *utilización*; además, que es uno de los más importantes. Tres aspectos clave identificados en esta relación fueron: las recomendaciones detalladas, las acciones concretas y el apoyo en la evidencia (Johnson, 2009).

iii) *Calidad de la evaluación*. Este factor, denominado por Cousins y Leithwood (1986) como *calidad de la evaluación*, se orienta a las características del proceso evaluativo, y es definido a partir de la sofisticación metodológica, el rigor, el tipo de modelo de evaluación empleado o la intensidad de las actividades evaluativas.

Más del 40 por ciento de los estudios analizados para el periodo 1971-1985 abordaron este factor y en ellos se encontró tanto relación como no relación con la *utilización* de las evaluaciones. Johnson (2009), por su parte, analizó seis estudios donde se mostró relación entre la *calidad* y la *utilización* pero, además, se encontró que ésta es menos importante que los hallazgos y la calidad de la comunicación.

En este caso, he decidido replantear el factor y explorar la evidencia empírica disponible a través de dos nuevas categorías: i) la *sofisticación y rigor metodológico*, y ii) la *intensidad de la evaluación*. Hago este ajuste porque considero que el nombre designado por Cousins y Leithwood (1986), “calidad de la evaluación”, no es adecuado, debido a que es poco preciso y a que los estudios empíricos no cubren la complejidad del término “calidad” al abordar solamente elementos referentes a la metodología y la intensidad de esta práctica.

Evidencia sobre la sofisticación y rigor metodológico. Los resultados de los estudios empíricos revisados por Cousins y Leithwood (1986) fueron poco concluyentes. Por una parte, Weeks (1979), Van de Vail y Bolas (1982), y Siegal y Tuckel (1985) reportaron que el aumento de la sofisticación metodológica inhibe la utilización de la evaluación. Asimismo, Yeh (1980) concluyó que las decisiones son más propensas a estar basadas en pruebas menos sofisticadas.

Por otro lado, Alkin et al. (1974), Granville (1977), y Patton et al. (1977), mostraron que la mejora en la sofisticación metodológica se relaciona positivamente con la *utilización* para la toma de decisiones y el desarrollo conceptual de los usuarios (citados en Cousins y Leithwood, 1986).

Dickey (1980) y Dickman (1981), por su parte, no encontraron ninguna relación entre este factor y la *utilización*; pero observaron que la consistencia metodológica estaba directamente relacionada con el cambio sobre la percepción del programa.

Evidencia sobre la intensidad de la evaluación. Esta variable se analiza solo en dos estudios que obtienen hallazgos confrontados. Barnette y Thompson (1979), por un lado, encontraron que la intensidad de las evaluaciones conduce a una mayor *utilización* de los resultados para la toma de decisiones. No obstante, Salmón-Cox (1981), concluye que el aumento en la frecuencia de las pruebas estandarizadas se traduce en una menor utilización para la toma de decisiones (citados en Cousins y Leithwood, 1986).

iv) *Credibilidad.* La credibilidad del evaluador o del proceso evaluativo es definido en términos de la objetividad, la verosimilitud, y la adecuación de las actividades del evaluador.

En el trabajo de Cousins y Leithwood (1986), más de una cuarta parte de los estudios analizados encontraron una relación entre la credibilidad y la *utilización* de las evaluaciones. Este factor fue analizado en función de la reputación (Alkin et al., 1979; Dawson y D'Amico, 1985), y de las particulares del evaluador (Braskamp et al., 1978; Dickman, 1981; Newman et al., 1979).

En el trabajo de Johnson (2009), cuatro estudios abordaron la credibilidad y obtuvieron conclusiones divididas en varios hallazgos: dos encontraron un fuerte vínculo

con la *utilización* a partir del conocimiento del evaluador sobre el programa (Barrios, 1986) y la reputación (Boyer and Langbein, 1991); y dos estudios no encontraron tal relación (Bober y Barlett, 2004; Johnson, 1993).

Los estudios empíricos, en gran medida, han explorado la credibilidad a partir de las particularidades del evaluador (la reputación, el conocimiento sobre el programa, la formación y el nivel de capacidades, entre otros). Al respecto, considero que las variables usadas no son las adecuadas para representar la compleja definición que hacen los autores sobre este factor. Por tanto, he decidido renombrarlo como “particularidades del evaluador”, y no abordar la credibilidad debido a la subjetividad del término y la complejidad que implicaría su medición.

- v) *Relevancia*. La relevancia se define a partir de la capacidad de las evaluaciones para responder a las necesidades de información de los tomadores de decisiones, considerando el propósito de la evaluación y el nivel en que se encuentran los usuarios dentro de la organización.

En el análisis realizado por Cousins y Leithwood (1986), 18 estudios exploraron la *relevancia* y encontraron un vínculo entre ésta y la *utilización*. Los estudios exploraron el grado en que las evaluaciones se orientaban a la audiencia o si el evaluador era interno o externo a la organización. Los evaluadores internos se consideraron con mayor conocimiento acerca de las características de las organizaciones.

La mayoría de los estudios concluyeron que las evaluaciones que reflejan conocimiento sobre el contexto donde serán aplicadas, se asocian a un nivel más alto de *utilización*: apelan a las preferencias de los tomadores de decisiones; buscan el consenso sobre el problema de la evaluación; o demuestran conocimiento sobre la operación del programa y la toma de decisiones (D’Amico, 1985; Osterlind 1979; Van de Vail y Bolas, 1982). Al contrario, las evaluaciones que ignoran estas cuestiones fueron asociadas a niveles relativamente bajos de *utilización* (David, 1978; Florio et al. 1979; McGowan, 1976; Neigher, 1979; Sproull y Zubrow, 1981; Wholey & White, 1973).

En el análisis de Johnson (2009), seis estudios abordaron la relevancia y obtuvieron conclusiones mixtas. Dos no encontraron vínculo entre la *relevancia* y la *utilización*; pero

otros dos sí encontraron relaciones fuertes entre la pertinencia de la información y su *utilización*.

vi) *Puntualidad*. Cousins y Leithwood (1986) definen la puntualidad en función de la pertinencia del momento en que se difunden los resultados de la evaluación a los tomadores de decisiones.

En la revisión de estudios empíricos llevada a cabo por Cousins y Leithwood, en menos del 15 por ciento de los casos se examinó la relación entre la *utilización* y la puntualidad con que se comunican los resultados. Los estudios arrojaron que la entrega oportuna de resultados se relaciona positivamente con la utilización (Dickey, 1980), las decisiones programáticas (David, 1978; Siegal y Tuckel, 1985; Wholey y White, 1973, 1973), y las decisiones políticas (Florio et al. 1979; Rich, 1979). No obstante, Patton et al. (1977), concluyó que la tardanza de los informes y la puntualidad tenían poco impacto.

En la investigación realizada por Johnson (2009), siete estudios abordaron la puntualidad y la mayoría encontraron una relación positiva entre la puntualidad y la *utilización*. Solo un estudio concluyó que la puntualidad no era relevante.

Los autores realizan un esfuerzo importante por armar un marco analítico en torno a los diversos estudios que se han llevado a cabo para explicar qué determina que una evaluación sea utilizada. No obstante, dicha propuesta establece definiciones ambiciosas (por ejemplo: “calidad de la evaluación” o “calidad de la comunicación”) que difícilmente son representadas por las variables que, en la práctica, son seleccionadas con limitaciones de tiempo, recursos e información.

De lo anterior se podría deducir que a pesar de que resulta tentador emplear la terminología que proponen los autores, esto sería inapropiado si consideramos que los criterios para seleccionar las variables son heterogéneas y que un mismo factor en estudios diferentes explora aspectos distintos. En otras palabras, que estos seis factores más que unidades de análisis son categorías que los autores establecen para facilitar su investigación sobre los estudios empíricos realizados. Esto ayuda a explicar el hecho de que no se obtuvieran resultados concluyentes sobre la relevancia de los seis factores.

En el primer apartado del capítulo 3, exploro la relevancia que tienen algunas características de las evaluaciones de Sedesol para su *utilización* a través de seis variables (número de recomendaciones, citas de internet en Google y Google *Scholar*, número de evaluaciones, procedencia del evaluador y pertinencia) que se vinculan a cinco de los factores mencionados (hallazgos, comunicación, intensidad de las evaluaciones, credibilidad y relevancia).¹²

¹² La *puntualidad* no se analiza debido a que no hay información pública disponible sobre las fechas de entrega de los informes de evaluación.

1.2.2. FACTORES INSTITUCIONALES: LA INDEPENDENCIA DEL PROCESO EVALUATIVO

En la bibliografía disponible, prácticamente todos los trabajos que abordan la *utilización* de las evaluaciones destacan, en mayor o menor medida, la relevancia del marco normativo. No obstante, Balthasar (2009), fue el único autor que identifiqué que explora la relación entre los factores institucionales y la *utilización* de los resultados. Factores que, en este caso, determinan la distancia entre evaluador y evaluado.

William Trochim (2009) es otro autor que realiza un esfuerzo por articular el marco normativo que influye en la evaluación gubernamental. Trochim propone un esquema a partir del cual se pueden identificar y analizar los arreglos institucionales que definen la política nacional de evaluación, que en muchos casos solo está implícita en la regulación.

En esta sección, abordo los criterios que Balthasar (2009) emplea en su investigación.

vii) Actor detonante de la evaluación: ¿la iniciativa se decide por una unidad independiente a los responsables de la evaluación?

Para abordar este factor, Balthasar (2009) analiza si los arreglos institucionales que definen al responsable de las iniciativas de evaluación determinan la capacidad de los evaluadores para hacer frente a las necesidades de información de los destinatarios. Encontrando en su investigación que hay más dificultades para hacer preguntas sobre el efecto y la pertinencia de las medidas revisadas cuando se designa la responsabilidad de la evaluación a actores próximos a la intervención pública. En este sentido, Mayne, Divorski y Lemaire (1998) concluyen que la facultad para iniciar una evaluación también incluye la facultad para especificar los temas a tratar.

A partir de sus hallazgos, Balthasar (2009) afirma que la distancia entre evaluadores y evaluados es menor si los últimos son los responsables de efectuar la evaluación. Por tanto, si un proyecto de evaluación se origina en una unidad que no está involucrada directamente en el proceso, se asume que existe una mayor distancia entre evaluadores y evaluados. Esto debido a que los evaluados no pueden interferir en la determinación de sí o cómo una evaluación en su área se llevará a cabo.

viii) Unidad responsable del proceso: ¿existe una unidad involucrada en la evaluación?

El segundo factor institucional observado por Balthasar es la existencia de una unidad responsable de las evaluaciones, a la que se le otorgan responsabilidades con respecto a la preparación, seguimiento y valoración de la práctica evaluativa (Balthasar, 2009).

El papel de esta unidad es mediar la comunicación entre los evaluados, la oficina y los evaluadores externos; establecer las especificaciones de rendimiento y para la contratación de las evaluaciones en coordinación con los responsables de los programas; y guiar el proceso evaluativo, desde la selección de los evaluadores externos hasta su ejecución (Balthasar, 2009).

El autor asume que la distancia entre evaluadores y evaluados es mayor si una unidad ajena al proceso es la encargada de iniciar la evaluación y supervisarla, que si los responsables del programa gestionan la evaluación por sí mismos.

ix) Asimilación de costes: ¿la evaluación es financiada con el presupuesto de la instancia evaluada?

El tercer factor institucional que analiza Balthasar (2009) para determinar la independencia del proceso evaluativo es la asimilación de los costes. En la práctica, se adoptan diversas medidas presupuestarias para financiar los recursos que demanda la evaluación de las intervenciones públicas, comúnmente éstos provienen de: los órganos de control, las agencias federales con presupuesto específico, o por las unidades responsables del proceso evaluativo.

En la revisión de literatura, Balthasar (2009) encuentra que existen muy pocos análisis sistemáticos sobre las repercusiones que pueden tener las diferentes formas de financiamiento para la evaluación. Una excepción es la obra de Wollmann (2001) que estudia las características de la investigación científica y el trabajo contratado. Por lo general, los aspectos financieros sólo son observados a través del costo de la evaluación (por ejemplo, Bussmann 1995, 56; Wottawa y Thierau 1998, 117; Tarjeta, Greeno, y Peterson 1992).

En el análisis de este factor, Balthasar (2009) asume que la distancia entre evaluadores y evaluados, y por tanto, la independencia del proceso, es mayor si los recursos financieros no provienen de la entidad evaluada.

- x) Asignación de contratos: ¿la implementación de las evaluaciones es asignada a través de una licitación?*

Para Balthasar (2009) la independencia del proceso evaluativo se determina, en parte, por los mecanismos de asignación de contratos. Desde este enfoque, la independencia es mayor si los contratos se designan a través de un proceso competitivo, es decir, si se emite una invitación abierta o selectiva a licitar.

Cuando se lleva a cabo una auto-evaluación, los directamente involucrados son comisionados para supervisar el proceso, por tanto, no hay distancia entre los evaluados y evaluadores. No obstante, si en estos casos los responsables de la evaluación trasladan el mandato directamente a un tercero, se puede suponer que existe una distancia — aunque pequeña — entre evaluadores y evaluados.

- xi) Implementación de las evaluaciones: ¿la evaluación es realizada fuera de la oficina?*

En la literatura hay repetidas referencias sobre la idea de que las evaluaciones independientes deben ser realizadas por personas que no están involucrados en la ejecución de las iniciativas (Landert 1996; Williams, de Laat, y Stern 2002, citado en Balthasar, 2009). Asimismo, para Balthasar (2009), la distancia entre los evaluadores y los evaluados, es grande si las investigaciones no se llevan a cabo por las personas responsables de la iniciativa.

Siguiendo el mismo criterio, se considera que la independencia del proceso es máxima si se lleva a cabo la evaluación al exterior de la administración (Müller-Kohlenberg y Beywl, 2003, p. 65; y Widmer, 2000, p. 80; citado en Balthasar, 2009).

xii) Publicación de resultados: ¿los resultados son puestos a disposición del público?

En el modelo de Balthasar (2009), se asume que la distancia entre evaluadores y evaluados es mayor si se publica el informe final. Para los evaluadores, este hecho significa que deben hacer su trabajo disponible para la discusión en la comunidad académica Wollmann (2003, p.344). Además, es la única manera de permitir meta-evaluaciones que tienen como propósito evaluar la calidad de las evaluaciones (p.ej. Widmer 1996).

En la discusión científica, la cuestión de si los resultados de las evaluaciones deben publicarse es de considerable importancia. Bussmann (1995; citado en Balthasar, 2009), al igual que otros autores (p.ej. Wollmann, 2003, p. 344), aboga por la máxima transparencia y la apertura en las evaluaciones, con la intención específica de permitir los procesos de aprendizaje. Valovirta (2002, p. 78), por su parte, hace hincapié en la función argumentativa de los informes de acceso público.

En cuanto a la difusión de los informes, Widmer, Landert y Bachmann (2000) consideran importante que todas las personas que están de alguna manera involucradas o son afectadas por la evaluación tengan acceso a los resultados. En los casos en que este grupo de personas es muy grande, el informe tendría que ser accesible al público (Widmer, Landert y Bachmann, 2000, p. 9).

1.3. ¿CUÁL ES EL EFECTO DE LA EVALUACIÓN EN LAS INTERVENCIONES PÚBLICAS?

En este apartado, exploro el desarrollo conceptual sobre la *influencia* de la evaluación. El objetivo es contar con un marco de análisis que me permita estudiar el efecto de las recomendaciones aplicadas sobre el mejoramiento de los programas valorados.

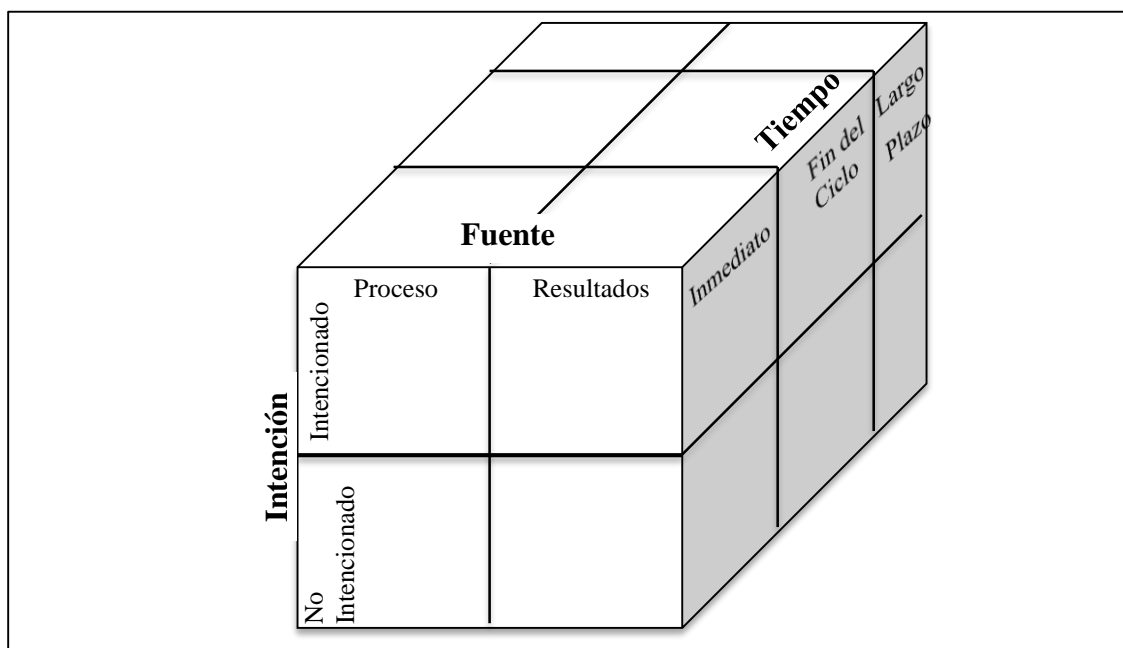
La *influencia* había sido estudiada, desde años atrás, como un tipo de uso de la evaluación; pero desde inicio del siglo XXI, aproximadamente, el estudio de la *influencia* ha propiciado el desarrollo de un nuevo marco de análisis a través del cual se explica el aprovechamiento de las evaluaciones, en términos todavía más amplios que la *utilización* (véase apartado anterior).

Desde este nuevo enfoque, se argumenta que los resultados pueden extenderse fuera del proceso evaluativo y que, a menudo, son transmitidos por profesionistas, académicos y evaluadores; impactando de manera indirecta a otros profesionistas o a expertos en políticas (Heclo, 1978, citado en Weiss, 1998). El conocimiento generado, durante y al final de las evaluaciones, es reconocido como una fuerza importante en la formulación de políticas y programas; y la acumulación de conocimiento de esta práctica como un medio de *influencia*. En este sentido, a través de la *influencia*, los resultados se infiltran en las coaliciones de defensa, alteran paradigmas de política, cambian la agenda política y afectan sistemas de creencias dentro de las instituciones. Este nivel de impacto no es común ni fácil pero la evaluación puede contribuir a cambios a gran escala en el pensamiento y, en última instancia, a cambios en la acción (Weiss, 1998, p. 24).

Existen tres trabajos de gran relevancia en el desarrollo conceptual de la *influencia*: Kirkhart (2000), Alkin y Taut (2003), Henry y Mark (2003 y 2004). Kirkhart (2000), la primera en emplear el término *influencia* (al que nos referimos en el resto de los capítulos como efecto), afirma que los términos “uso” y “utilización” no son adecuados porque sugieren una acción intencionada, mientras que en la realidad, los efectos de la evaluación son más difusos (Kirkhart, 2000; citado en Ledermann, 2012, p. 161). La autora sugiere remplazar los términos *uso* o *utilización* por el de *influencia*, definiendo este último como “la capacidad o poder de personas o cosas de producir efectos en otros mediante medios intangibles o indirectos” (2000, p. 7; citado en Johnson et al., 2009, p. 378).

Kirkhart (2000) elabora el primer modelo conceptual sobre la *influencia*; en él sugiere el análisis de los efectos de la evaluación a través de tres dimensiones (véase *ilustración 1.3*): la intencionalidad del evaluador respecto al efecto, la fuente de la influencia (proceso o hallazgos) y, el tiempo que tarda en manifestarse la influencia (inmediato, al concluir la evaluación, o a largo plazo). Este modelo no solo sentó las bases para el análisis de la *influencia* sino que hoy en día sigue siendo la referencia principal en esta línea de investigación, en gran medida, los nuevos desarrollos conceptuales parten de esta propuesta.

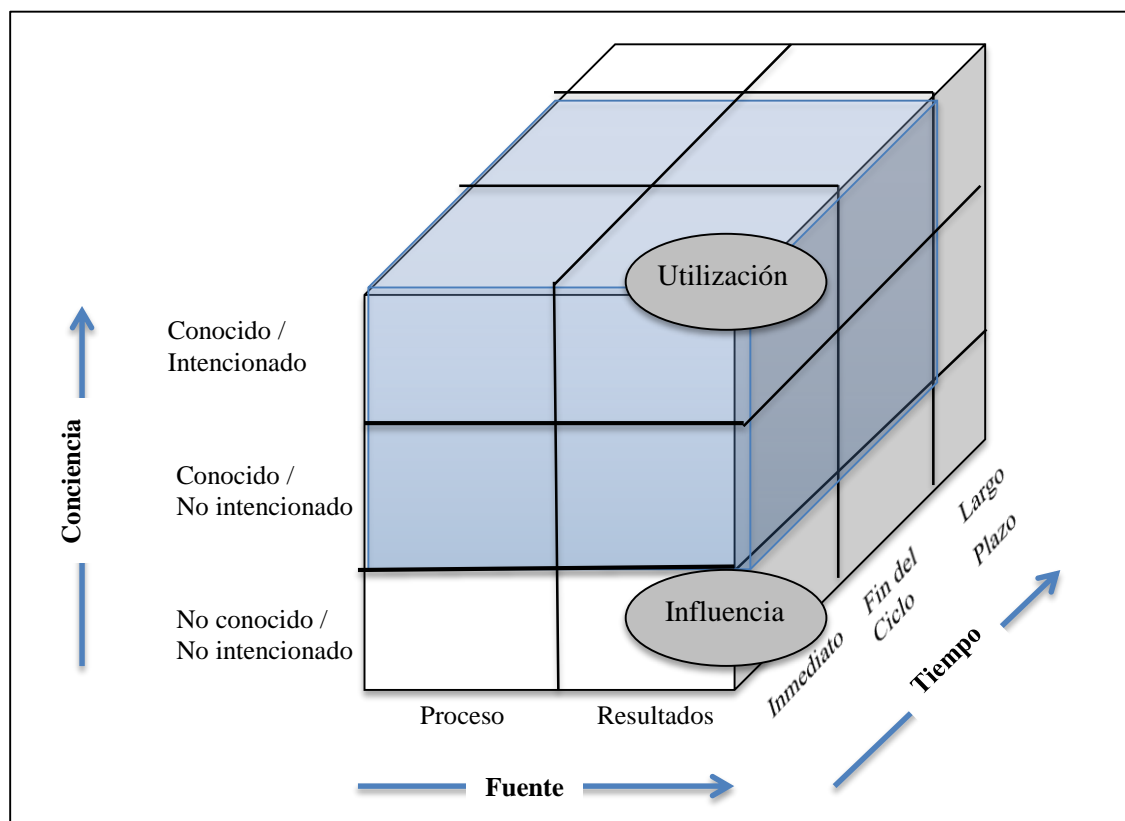
Ilustración 1. 3. Marco de análisis de Karen Kirkhart sobre la Influencia (2000)



Fuente: Kirkhart, 2000.

Alkin y Taut (2003), retoman el primer modelo de Kirkhart y agregan a su propuesta la consideración de los efectos no conocidos (véase *ilustración 1.4*). Estos autores no sugieren la eliminación del término *utilización* sino que lo diferencian de la *influencia*. Para Alkin y Taut, con la *utilización* los evaluadores están conscientes de los efectos generados, ya sean intencionados o no intencionados; y el concepto de *influencia* enriquece al de *utilización* al considerar que la práctica evaluativa también tiene impactos no conocidos y no intencionados por parte de los actores vinculados (2003: 9; citado en Johnson et al., 2009, p. 378).

Ilustración 1. 4 Nuevo modelo sobre la Influencia propuesto por Alkin y Taut (2003), adaptando el realizado por Kirkhart (2000)



Fuente: Alkin y Taut, 2003.

Henry y Mark (2003) también toman como referencia el trabajo de Kirkhart para crear un modelo donde identifican mecanismos a través de los cuales se puede observar el efecto de la evaluación. Los mecanismos son clasificados en cuatro grandes procesos (generales, cognitivos, motivacionales y comportamiento) y tres niveles de análisis (individual, interpersonal y colectivo) (véase cuadro 1.2). Un año más tarde, estos autores refinan su modelo y enriquecen la investigación evaluativa vinculando teorías de otras disciplinas con el objetivo de ayudar a describir cómo el proceso evaluativo genera cambios.

Mark y Henry (2004), buscan ampliar la manera en que los evaluadores conceptualizan las consecuencias de su trabajo; argumentan que el objetivo de la evaluación es el mejoramiento social, motivo por el cual, sugieren la necesidad de identificar a través de diferentes caminos y niveles los mecanismo mediante los cuales las evaluaciones llevan este último objetivo. Así, los autores se centran en las consecuencias de la evaluación que están relacionadas al mejoramiento de las condiciones sociales.

Cuadro 1. 2 Mecanismos que pueden mediar la influencia de la evaluación, Mark y Henry, 2003

Tipo de proceso	Nivel de Análisis		
	Individual	Interpersonal	Colectivo
Influencia general	Elaboración Heurística Obscurecimiento La adquisición de habilidades	Justificación Persuasión Agente de cambio Influencia opinión de minoría	Ritualismo Audiencia legislativa La formación coaliciones Proyecto de legislación Configuración de estándares Consideración de política
Cognitivo/Afectivo	Relevancia Opinión/actitud de valencia	Normas locales de referencia	Configuración de la agenda Políticas orientadas al aprendizaje
Motivacional	Objetivos personales y aspiraciones	Normas cautelares Recompensa social Intercambio	Incentivos estructurales Fuerzas del mercado
Comportamiento	Desempeño de nueva habilidad Cambio individual en la práctica	Cambio de colaboración en la práctica	Continuación, cese o cambio del programa Cambio de política Difusión

Fuente: Mark y Henry, 2003.

Algunos trabajos han empleado los modelos desarrollados por Henry y Mark (2003 y 2004); sin embargo, poco se ha avanzado en la conceptualización de los mecanismos que se señalan.

Herbert (2014), llevó a cabo una revisión de la literatura disponible donde el término influencia es usado para investigar el impacto de las evaluaciones. A partir del análisis de veintiocho estudios, en los cuales se emplea una variedad de definiciones y métodos, Herbert encontró que la manera en cómo el análisis de la *influencia* era incorporada en las investigaciones difería considerablemente. Algunos de los estudios descriptivos incluyeron el marco de análisis de Kirkhart (2000) para abordar los impactos de la evaluación interpretando la *influencia* como los efectos intencionados y no intencionados de la evaluación en un largo plazo; mientras que otros lo emplearon simplemente para enmarcar investigaciones sobre el impacto de la evaluación, en términos amplios.

También encontró que había otro grupo de investigaciones orientadas a identificar los factores vinculados a la *influencia* de la evaluación, basándose en el marco de análisis propuesto por Mark y Henry (2004). En los estudios más sencillos, solo se analiza los factores determinantes de la *influencia* sin pretender entender las circunstancias en que ésta se genera (McEathron, 2008; y Lehtonen, 2010).

En esta investigación, a pesar de la diversidad de mecanismos que sugieren Henry y Mark (2003), analizo exclusivamente efectos intencionados (tanto inmediatos como no inmediatos) que derivados de los resultados de la evaluación se manifiestan como cambios a los programas (señalados por los autores como mecanismo de comportamiento colectivo).

1.4. EL APROVECHAMIENTO DE LAS EVALUACIONES EN MÉXICO

En el caso mexicano, al igual que en la mayoría de las iniciativas nacionales de evaluación, el *aprovechamiento* de los resultados enfrenta como principal riesgo su burocratización. En otras palabras, que la evaluación se norme, se programe, se contrate y realice con estándares técnicos aceptables, pero que no se utilice para diseñar políticas sociales a mejorar las condiciones de vida de la población mediante una operación eficaz (Cardozo, 2009).

Al respecto, Maldonado (2011, p. 228) señala que a pesar de los avances logrados, se sabe poco sobre el grado de apropiación de los hallazgos en los procesos internos de gestión, de la calidad y de la pertinencia de las recomendaciones derivadas del proceso evaluativo; así como de los factores externos que influyen en la capacidad de los operadores para detonar procesos internos de mejora, y sobre todo, de su contribución específica al desempeño de los programas.

Los pocos estudios que abordan el *aprovechamiento* de las evaluaciones en México se centran, principalmente, en identificar los retos que dificultan que las recomendaciones se utilicen (González, Zabaleta y Cerdán, 2008; Cardozo, 2009; Cejudo, 2011; Montiel, 2011; Patiño, 2014). No obstante, en algunos de estos trabajos ya se están realizando esfuerzos por observar el efecto de esta práctica en los programas (Cejudo y Maldonado, 2011; Gris, 2014; Pérez, 2015).

En este apartado, presento brevemente los aspectos más sobresalientes en la literatura sobre la *utilización* de las evaluaciones y sobre los esfuerzos realizados para analizar su *influencia*. Como se podrá observar, en mayor o menor medida, gran parte de los trabajos abordan ambos temas.

1.4.1. ANÁLISIS DE FACTORES QUE EXPLICAN LA UTILIZACIÓN

En la literatura disponible para el caso mexicano, encontré que la *utilización* de las evaluaciones es un tema reciente y poco explorado. Sin embargo, identifiqué algunas investigaciones que, con análisis cualitativo, exploran los factores más relevantes para su aprovechamiento, destacando las siguientes: Myriam Cardozo (2009), resaltó un estudio sobre la institucionalización de la cultura de evaluación en México identifican los principales problemas que enfrenta la implementación de los resultados. Cejudo y Maldonado (2011), coordinan una serie de estudios encabezados por el CIDE (con el apoyo del gobierno federal) y examinan diez casos de “buenas prácticas” para valorar el compromiso de los programas con la implementación de las evaluaciones. Patiño (2014) analiza la aplicación de las recomendaciones en la toma de decisiones para cinco programas de Sedesol que atienden grupos vulnerables. Gris (2014) estudia la aplicación de las recomendaciones para estimar la utilización de las evaluaciones como herramienta de innovación gubernamental.

Siguiendo el debate internacional sobre la *utilización*, estas investigaciones ayudan a identificar factores que permiten explicar por qué unas evaluaciones se utilizan más que otras (véase *ilustración 1.5*).

En los primeros años de la práctica evaluativa en México, Cardozo (2009) identificó cuatro aspectos que inhibían la utilización de los resultados: un mercado de evaluación con oferta restringida y problemas de formación (principalmente entre funcionarios vinculados y los diputados que deben emplear los resultados para asignar el presupuesto), la emisión de pocos juicios explícitos de valoración fundamentados del programa analizado, poca difusión de los resultados, y un limitado avance en la institucionalización de la cultura evaluativa, por lo que prevalece la idea sobre esta práctica como un requisito administrativo más que una actividad de aprendizaje y mejoramiento de la gestión.

Posteriormente, Cejudo (2011) explora la relevancia de los atributos de la evaluación¹³ y de las organizaciones¹⁴ para la *utilización* de los resultados, y concluye que no todos son

¹³ Los factores referentes a los atributos de la evaluación son: la relevancia, la credibilidad, la calidad, los hallazgos mismo, las prácticas de comunicación del evaluador y la puntualidad de los reportes (Cejudo, 2011).

relevantes para identificar posibles obstáculos en el caso mexicano. Al respecto, el autor resalta dos particularidades de la evaluación nacional: un sistema estandarizado que no presenta muchas variaciones en cuanto a tiempos y formatos y un entorno institucional que presta más atención al cumplimiento de los procesos administrativos que a la generación de valor público. Esto último genera que los evaluadores y los operadores de los programas tengan más interés en que las evaluaciones arrojen datos positivos que en obtener lecciones útiles para mejorar el desempeño del programa.

De acuerdo con Cejudo (2011), los desafíos que enfrenta la evaluación en México para convertir los resultados en acciones de mejora son cuatro: Primero, la pertinencia de los hallazgos, es decir, que estén conectados con la lógica causal de la política pública, y no sólo aludan a aspectos secundarios que no sean relevantes para los propósitos de la política. Segundo, la factibilidad de las recomendaciones, de acuerdo con las capacidades administrativas, legales, organizativas y presupuestales de los usuarios. Tercero, la calidad y precisión de las recomendaciones específicas al programa deben ser puntales para definir no sólo los propósitos sino los instrumentos que podrían emplearse. Cuarto, claridad sobre quiénes son los responsables de atender las recomendaciones, lo que implica tener conocimiento acerca de la operación del programa y de la organización pública de su implementación.

A través de entrevistas, Montiel (2011)¹⁵ encuentra que, en principio, solo cumplen aquello que está claramente normado y que el diseño institucional actual resulta aún limitado en cuanto al aprovechamiento de los resultados ya que en el “mecanismo de seguimiento” no se prevé un agente que verifique qué ajustes se hacen a los programas, si su cumplimiento implica aumentar los costos de su operación, o quiénes son los responsables de hacer las modificaciones. Además, aparte de las reglas formales (LGDS, LFPRH, Lineamientos, PAE, Mecanismo de Seguimiento) y los incentivos planteados en el diseño institucional, hay otros factores que favorecen que las evaluaciones se utilicen en la fase de implementación: el liderazgo,¹⁶ la calidad de la evaluación,¹⁷ la claridad de las reglas y su obligatoriedad y, las

¹⁴ Los factores referentes a los atributos organizacionales, identificados en la literatura son: el compromiso, las necesidades de información, la información en competencia, las características personales, el clima organizacional, el clima político y el clima financiero (Cejudo, 2011).

¹⁵ El trabajo de María José Montiel forma parte de la investigación del CIDE que coordinaron Cejudo y Maldonado en 2011.

¹⁶ Al margen del cumplimiento de los requerimientos administrativos, Montiel identificó que la conciencia que permea desde los mandos altos sobre la importancia y la utilidad de la evaluación significa un factor muy importante.

limitaciones de tiempo.¹⁸

Montiel también concluye que como el Coneval comparte atribuciones (con la SHCP y la SFP) su fuerza legal es menor; que el presupuesto con el que opera y el personal asignado son insuficientes (bajo en proporción con el resto de las entidades reguladoras) y; que tiene escasas atribuciones para sancionar el incumplimiento de las dependencias o entidades valoradas. Aspectos que de acuerdo con la autora, denotan una baja jerarquía institucional.

Maldonado (2011), por su parte, enfatiza la relevancia de cuatro aspectos: la pertinencia de las recomendaciones, que en ocasiones ofrecen pistas contradictorias e inconsistentes para la mejora de los programas; el déficit de capital humano, relacionado con una gran heterogeneidad y un nivel relativamente bajo de capacidades en materia de evaluación; la asignación clara de responsabilidades y la capacidad de trabajo en equipo coordinado.

A partir de entrevistas a funcionarios clave, Patiño (2014) identifica que, en cinco programas de Sedesol orientados a población vulnerable,¹⁹ la implementación de las recomendaciones está determinada por algunos factores organizacionales como: el liderazgo y el compromiso de las unidades evaluadas; la capacidad técnica de los programas para utilizar la información de las evaluaciones (posibilidad de generar información útil para la evaluación y transformar los hallazgos en acciones puntuales) y la factibilidad de las recomendaciones (condicionada por la percepción de los usuarios y el entendimiento sobre el programa que provee la colaboración entre evaluador y evaluado). En cuanto a su aprovechamiento para ajustes presupuestales, Patiño concluye que los aspectos más relevantes son: las decisiones políticas y el diseño institucional.

Gris (2014), en su análisis sobre la implementación de las recomendaciones y el cambio del programa Oportunidades, enfatiza la relevancia de las características organizacionales en la capacidad innovadora de la evaluación y señala que para incorporar los resultados en los

¹⁷ La autora define la calidad de la evaluación en torno a la capacidad de los resultados de reflejar la información proporcionada por las áreas responsables, su apego a la realidad del programa valorado y, principalmente, si la unidad responsable percibe que se ha hecho un trabajo profesional (Montiel, 2011, p. 57).

¹⁸ Se identificó que los plazos definidos para la elaboración y el seguimiento de las evaluaciones (un ciclo presupuestario) generan círculos burocráticos que disminuyen las ventajas de la evaluación (Montiel, 2011).

¹⁹ Programa Atención a Jornaleros Agrícolas (PAJA); Coinversión Social (PCS); 70 y Más; Apoyo a las Instancias de Mujeres en las Entidades Federativas, para implementar y ejecutar programas de prevención de la violencia contra las mujeres (PAIMEF) y; el programa de Fortalecimiento a la transversalidad de la perspectiva de género.

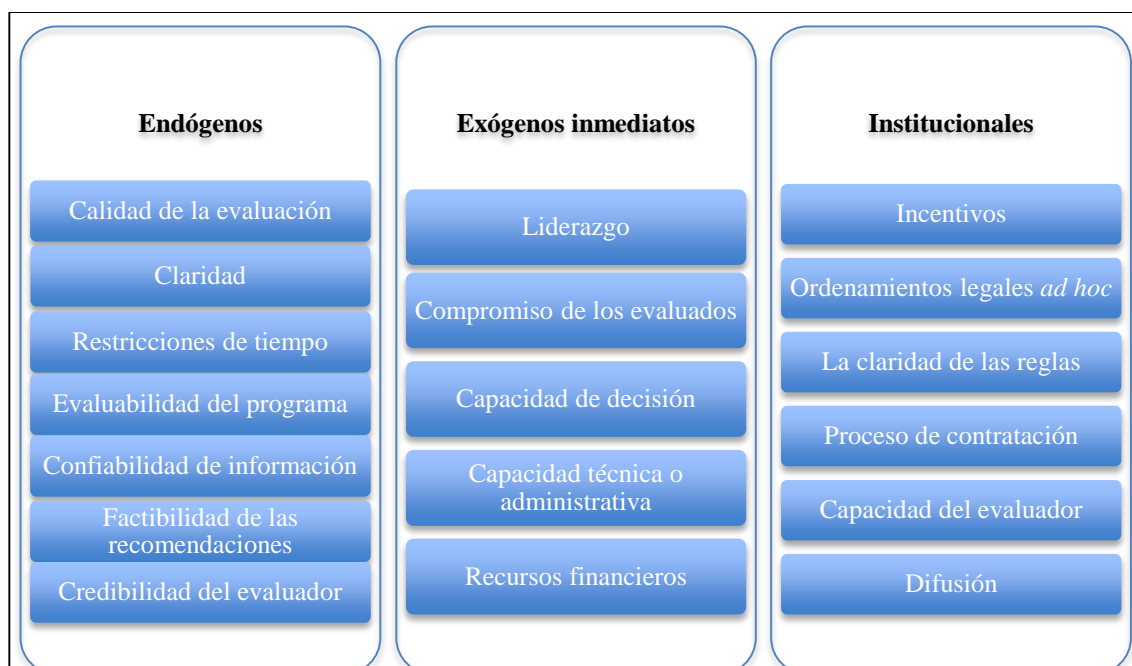
procesos de toma de decisiones, además del contexto político, son pertinentes: la promoción de las evaluaciones o de los estudios con resultados relevantes y oportunos; la capacidad de analizar los hallazgos y traducirlos en planes de trabajo susceptibles de implementación y la capacidad de incorporar los hallazgos en discusiones más amplias sobre la efectividad de la política.

A partir de su investigación, Carolina Gris observa un cumplimiento oportuno en el seguimiento formal de las recomendaciones, pero identifica como principal reto de la *utilización* que ésta no se restrinja a las áreas operativas de los programas sino que organizaciones de la sociedad civil y partidos políticos (entre otros), también implementen las evaluaciones para participar con mayores elementos en el debate público acerca de la efectividad gubernamental.

Cabe mencionar que la práctica evaluativa en torno al programa Oportunidades es excepcional, ya que desde su inicio ha estado supervisada por organismos internacionales y es un ejemplo de “éxito” para México.

Por último, Coneval, que lleva a cabo una revisión de la política nacional de evaluación por motivo de su décimo aniversario; afirma, a partir de la experiencia general con Sedesol, que la evaluación y la medición de resultados han sido necesarias en una sociedad cada vez más democrática y transparente, pero no condición suficiente para reducir los problemas de pobreza y la desigualdad en México. Por ello, señala la necesidad de fortalecer su utilización, a partir de una mayor vinculación entre el proceso de medición de los problemas nacionales, la evaluación y la conformación del presupuesto (Coneval, 2016).

Ilustración 1. 5 Factores determinantes en la implementación de los resultados de las evaluaciones



Fuente: Elaboración propia con datos de Cardozo, 2009; Cejudo y Maldonado, 2011; Patiño, 2014; Gris, 2014.

Cuadro 1. 3 Retos en la implementación de los resultados de las evaluaciones

Cardozo 2009	<ul style="list-style-type: none"> - Oferta de evaluadores restringida y con problemas de formación - Pocos juicios de valor explícitos y fundamentados - Poca difusión de resultados - Cultura de evaluación enfocada en lo administrativo
Cejudo 2011	<ul style="list-style-type: none"> - Pertinencia de los hallazgos - Factibilidad de las recomendaciones - Calidad y precisión de las recomendaciones - Claridad sobre las responsabilidades en la atención de las evaluaciones
Maldonado 2011	<ul style="list-style-type: none"> - Pertinencia de las recomendaciones - Capital humano - Trabajo en equipo
Montiel 2011	<ul style="list-style-type: none"> - Baja jerarquía institucional del Coneval
Patiño 2014	<p><i>Mejoramiento de programa:</i></p> <ul style="list-style-type: none"> - Mejora del Mecanismo de seguimiento - Capacitación de funcionarios - Creación de cuerpos colegiados para la calidad de los compromisos - Desigualdades en compromiso y utilización de recomendaciones - Reforzar el Modelo Sintético de Desempeño - Ajustar calendario de PAE y Mecanismo de seguimiento con proceso presupuestal <p><i>Ajustes presupuestales:</i></p> <ul style="list-style-type: none"> - Uso simbólico de las evaluaciones - Mejorar mecanismos e incentivos
Coneval 2016	<ul style="list-style-type: none"> - Mayor vínculo entre la evaluación, los problemas y el presupuesto

Fuente: Cardozo, 2009; Cejudo y Maldonado, 2011; Patiño, 2014; CONEVAL, 2016.

1.4.2. ANÁLISIS DEL EFECTO DE LAS RECOMENDACIONES

Como mencionamos previamente, algunas de las investigaciones que abordaron la utilización de las evaluaciones también aportaron evidencia empírica sobre su efecto en los programas: Cejudo y Maldonado (2011) comparten la experiencia de programas que implementaron exitosamente acciones de mejora para incrementar su desempeño; Patiño (2014) observa el efecto de las recomendaciones en la toma de decisiones de cinco programas de Sedesol; Gris (2015) explora la utilidad de los ASM comprometidos por el programa Oportunidades como medio de legitimación política; Pérez (2015), estudia la influencia de la evaluación externa en Programas para Educación Básica; y, Coneval (2016), a partir de las recomendaciones atendidas, identifica cinco tipos de mejoras implementadas a los programas (corregir actividades o procesos, modificar apoyos, reorientarlo sustancialmente, adicionar elementos o reubicarlo, o en su caso, suspenderlo).

El trabajo del CIDE, coordinado por Cejudo y Maldonado (2011), se realiza con el ánimo de incentivar a las dependencias y entidades a utilizar las evaluaciones para mejorar la operación de los programas o el uso de recursos. Para ello, en colaboración con la Secretaría de la Función Pública (SFP) y la asociación civil México Evalúa, se convocó al “Premio 2011” para Programas Federales Comprometidos con el Proceso de Evaluación.

Para esta premiación, se analizó si los hallazgos seleccionados eran factibles, así como su potencial para mejorar los resultados; además, si las acciones de mejora derivadas favorecían los procesos de gestión y el desempeño del programa. La principal aportación de esta actividad fue comprobar que la evaluación logra su cometido, con base en el compromiso decidido y la voluntad de cambios de los responsables de los programas. Otro hallazgo relevante fue que “las evaluaciones son sin duda tomadas en consideración por los programas, y sus recomendaciones, cuando son utilizadas, tienen gran potencial para mejorar el desempeño de los programas” (Cejudo y Maldonado, 2011).

Patiño (2014), por otra parte, analiza la aplicación de las recomendaciones en la toma de decisiones como insumo para mejorar los programas o para la designación del presupuesto. La autora analiza cambios en las reglas de operación de los programas y reacomodos en la estructura programática presupuestal (fusión, eliminación o creación de nuevos programas); y

concluye que los programas utilizan las recomendaciones y hallazgos de las evaluaciones porque están obligados por la normatividad, pero que en las decisiones presupuestales solo existe un uso simbólico por motivaciones políticas.

Gris (2014) explora la evaluación del programa Oportunidades como herramienta para la innovación gubernamental y encuentra que además de la utilización de las evaluaciones como mecanismo para la toma de decisiones, observado a través de cambios a reglas de operación; el programa aprovechó los resultados como mecanismo de legitimación política, innovando y trascendiendo el ámbito administrativo para alcanzar un apoyo político que le permita su continuidad y permanencia. Apoyo que se ha sustentado en la evidencia generada, la promoción del programa como paradigma dentro de la política social, y la participación de organismos internacionales como el Banco Mundial y el BID.

Pérez (2015) enfatiza que existen pocos estudios empíricos sobre la influencia de las evaluaciones en la mejora de los instrumentos de política pública. Por ello, además de proporcionar evidencia sobre cambios experimentados, en su trabajo aplica un modelo donde analiza la influencia de la evaluación considerando factores interno y externos. Gabriela Pérez concluye que la influencia es limitada en el desempeño de los programas, y que en temas relevantes del diseño (objetivo, población, criterios y beneficios), ésta es afectada por la forma en que se instrumenta (tipo de evaluación) y el contexto (tamaño del programa, entorno político, y presencia de elecciones).

Por su parte, Coneval (2016) analizó los tipos de mejora llevadas a cabo en los programas de desarrollo social para el periodo 2010-2014, y estimó el porcentaje de recomendaciones que se emplearon para cinco categorías de análisis: corregir actividades o procesos, modificar tipos de apoyos brindados, reorientar sustancialmente el programa, adicionar elementos al programa o reubicarlo, y suspender totalmente las actividades del programa.

Atendiendo estas categorías, la evidencia muestra que durante el periodo analizado, las recomendaciones implementadas se orientaron principalmente a la corrección de actividades (17 a 27 por ciento) y en la reorientación de los programas (de 63 a 65.1 por ciento). Por otro lado, su efecto en la modificación de apoyos se redujo de 20 a 6.3 por ciento, mientras que su aplicación para la adición o reubicación de programas pasó de 4 por ciento en 2011 a solo 1.6

por ciento en 2014. También se observó que en ningún momento las evaluaciones han derivado en la suspensión de programas.

REFLEXIONES FINALES

Las propuestas teóricas desarrolladas sobre el *aprovechamiento* de la evaluación se centran, en las primeras décadas, hacia la aplicación directa de los resultados (su *uso* o *utilización*). Sin embargo, en estudios más recientes se enfatiza la necesidad de conocer los beneficios que la práctica evaluativa genera tanto para la gestión pública como para la sociedad (su *influencia*).

Si bien el análisis sobre los efectos de la evaluación es más reciente, considero que éste adquiere especialmente relevancia en aquellos casos donde a pesar de que la cultura de evaluación se ha institucionalizado y se observa un ejercicio sistemático de esta práctica durante un tiempo considerable (como en el caso de Sedesol), no se percibe su beneficio en la calidad de vida de la población. Lo que dificulta justificar los recursos económicos y humanos que se destinan a este fin.

En México, los análisis sobre el *aprovechamiento* de las evaluaciones han consistido principalmente en estudios empíricos cualitativos que exploran la *utilización*. Y más recientemente, su efecto directo e intencionado en los programas (cambios en su diseño o implementación). En este sentido, la aportación principal de esta investigación, fue un estudio cuantitativo que me permitió estimar qué tanto se han *utilizado* las evaluaciones y, realizar un esfuerzo por determinar en qué medida su *efecto* ha ayudado a mejorar los programas (en el desempeño y la focalización de la población objetivo).

Algunas de las limitaciones del marco de análisis que empleo en esta tesis son: que no existe un sustento teórico que articule entre sí los factores endógenos estudiados, solo se observa la relevancia que tiene cada uno en la *utilización*; y que analizo solo los factores institucionales considerados por Baltasar (2009), por lo que omito otros que pudieran ser relevantes.

Por último, cabe destacar que en los estudios empíricos no se observa un seguimiento a todo el proceso del *aprovechamiento*. En general, se analiza de forma aislada el uso, la utilización o el efecto pero en ninguno de los casos se explora el vínculo entre la implementación de las recomendaciones y sus efectos a mediano o largo plazo. En este

trabajo, se investiga qué factores favorecen la *utilización*, qué cambios genera y si estos cambios han mejorado el desempeño de los programas o la consecución de sus objetivos.

CAPÍTULO 2

PANORAMA INICIAL SOBRE LAS EVALUACIONES REALIZADAS A LOS PROGRAMAS DE SEDESOL Y SU UTILIZACIÓN, 2007-2014

CAPÍTULO 2. PANORAMA INICIAL SOBRE LAS EVALUACIONES REALIZADAS A LOS PROGRAMAS DE SEDESOL Y SU UTILIZACIÓN, 2007-2014

En este capítulo presento un análisis introductorio sobre las evaluaciones que empleo como materia de análisis en el resto de la tesis. El objetivo es lograr un mayor entendimiento sobre cómo se ha dado la práctica valorativa en Sedesol y qué tanto se han utilizado los resultados obtenidos durante el periodo.

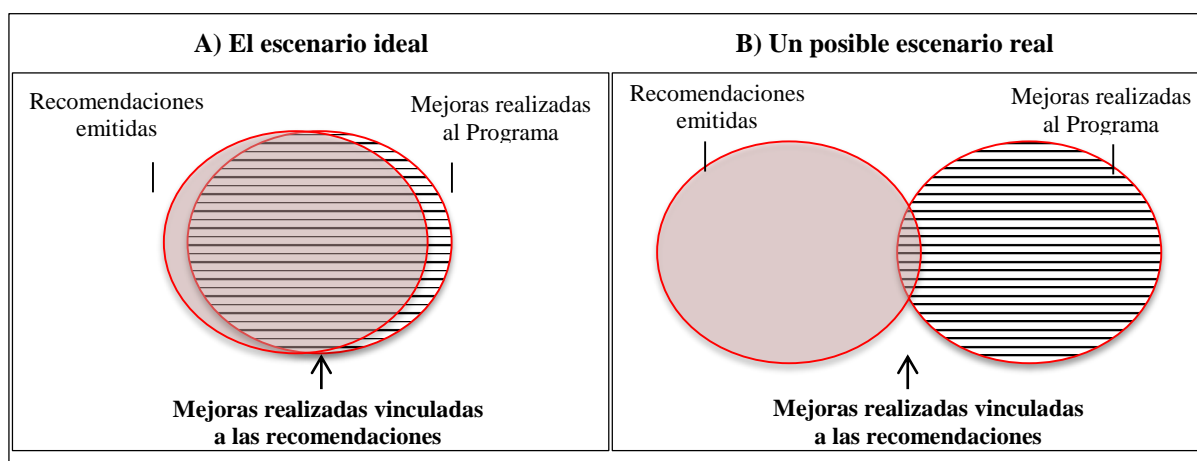
La relevancia de este análisis radica en que comúnmente se asume que las evaluaciones generan efectos tangibles en la implementación de los programas sin que exista, en la mayoría de los casos, evidencia que confirme la aplicación de las recomendaciones.

En este sentido, con el propósito de brindar una idea realista sobre el aprovechamiento de los resultados que tuvo Sedesol entre 2007 y 2014, exploro la utilización de 111 evaluaciones externas realizadas a programas de esta dependencia. Se trata de evaluaciones de diseño y desempeño coordinadas por Coneval y para las cuales la organización de la sociedad civil Gestión Social y Cooperación, A.C. (GESOC) calculó dos índices, uno que estima la calidad del diseño y otro la capacidad de los programas para lograr sus objetivos (*véase anexos*).²⁰

En la *Ilustración 2.1* se muestran dos posibles escenarios para la utilización de las evaluaciones. En el lado izquierdo, el escenario ideal, representa el aprovechamiento óptimo al que deberían aspirar todos aquellos que adoptan la práctica evaluativa; mientras que en la derecha se muestra uno de los posibles niveles de aprovechamiento que se pueden presentar en un escenario real expuesto tanto a la utilización idónea como a la no utilización.

²⁰ Estos índices presentan algunas limitaciones metodológicas que se analizan más adelante (*véase anexo*).

Ilustración 2. 1 Posibles escenarios sobre la utilización de la práctica evaluativa



Fuente: Elaboración propia.

A partir de esta investigación, identifiqué que las dependencias y entidades vinculadas a la práctica evaluativa de Sedesol han estudiado poco la información que generan e, incluso, que solo la han sistematizado de manera parcial.²¹ Por tanto, en este capítulo, decidí explorar parte de esta información para enriquecer el conocimiento disponible, principalmente entre actores que no participan directamente en el proceso evaluativo.

El capítulo se divide en dos apartados que contemplan la práctica y la utilización de las evaluaciones. En el primero de ellos brindo algunos antecedentes, exploro quiénes realizaron las valoraciones y cómo se financiaron (aspectos que podrían ser determinantes para la calidad, la credibilidad y la independencia de las evaluaciones). Estos factores, a su vez, podrían ayudar a explicar en qué medida se utilizan sus recomendaciones (*véase capítulo 3*).

En el segundo apartado, como ya mencioné, analizo la utilización de las evaluaciones. Para ello, documenté en qué medida las recomendaciones se emplearon para mejorar los programas de Sedesol a lo largo de los años. Exploro, en particular, cómo se utilizan las evaluaciones en dos etapas: primero, el porcentaje de recomendaciones convertido en Aspectos Susceptibles de Mejora (ASM) y, segundo, la proporción de estos aspectos que efectivamente se implementaron.

²¹ Emplee información de acceso público disponible en los portales oficiales de internet de las Secretarías vinculada al proceso de evaluación: Sedesol, CONEVAL, SHCP, SFP.

2.1. LA PRÁCTICA DE EVALUACIÓN EN SEDESOL

El gobierno mexicano ha implementado diversas iniciativas de evaluación desde la década de los setenta con el objetivo de optimizar el ejercicio de los recursos públicos y de legitimar decisiones gubernamentales, principalmente durante las crisis económicas que ha enfrentado el país. No obstante, la evaluación rigurosa y sistemática de la política social en México es reciente.

En los noventa, con un enfoque de gestión basada en resultados, Sedesol inició la evaluación del programa Oportunidades. Por la cual, el Banco Mundial (2009) reconoció a México como un pionero en América Latina en el uso de metodologías robustas para evaluar el desempeño de la política social.²²

Sedesol fue el principal promotor de Sistema de Evaluación y Monitoreo (SMvE) en el país, y realizó esfuerzos substanciales para fortalecer sus capacidades e instrumentos a favor de una política social dirigida al logro de resultados (Franco y Fernández, 2010). Los esfuerzos de esta Secretaría significaron cambios importantes en disposiciones legales y a un número creciente de evaluaciones en los programas sociales más relevantes (Banco Mundial, 2009).²³

En 2004, esta iniciativa sectorial de evaluación que encabezaba Sedesol originó una estrategia federal que se estableció en la Ley General de Desarrollo Social. Esta Ley creó el Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) y le otorgó la responsabilidad de mejorar la eficacia, la eficiencia, y la rendición de cuentas de la política de desarrollo social mediante la evaluación, el monitoreo de programas y políticas sociales.

²² En 1997, México recibe el apoyo del Banco Mundial para evaluar el Programa de Desarrollo Humano Oportunidades. Éste comenzó a operar en áreas rurales bajo el nombre de PROGRESA con el objetivo de mejorar la educación, la salud y la ingesta alimentaria de la población más pobre del país (Banco Mundial, 2008).

²³ Entre sus esfuerzos legales se encuentra la Ley General de Desarrollo Social, la Ley Federal de Presupuesto y Responsabilidad Hacendaria, y las normativas del Programa de Mejoramiento de la Gestión. Estas nuevas disposiciones legales proporcionan un marco que corrige, completa y refuerza normativas de evaluación preexistente, así como la sustentabilidad del sistema en el largo plazo (Banco Mundial, 2009).

El Coneval entró en operación en 2006 para conducir la valoración de los programas sociales y, desde entonces, ha consolidado avances en la cultura nacional de evaluación. Este organismo fue uno de los catalizadores en la institucionalización de la práctica evaluativa en el país, y sus características (independencia y capacidad técnica) le han permitido ejercer una gran influencia en el desarrollo de la gestión basada en resultados en el gobierno federal (Banco Mundial, 2009).

En este apartado observo las tendencias que ha seguido la práctica evaluativa de Sedesol desde que entró en operación Coneval: qué programas fueron seleccionados, quiénes realizaron las valoraciones, cuánto costaron y quién las financió; aspectos que suelen ser determinantes para la calidad, la credibilidad y la independencia de las evaluaciones.

Primero, menciono brevemente los programas de Sedesol que fueron evaluados entre 2007 y 2014, así como los criterios empleados para su selección. A continuación, exploro la participación que tuvieron organismos evaluadores y expertos en la materia. Esto es relevante tomando en consideración que la calidad y la credibilidad de los resultados están sujetos, en gran medida, a la capacidad que tienen los diferentes evaluadores para llevar a cabo su trabajo y al prestigio de los organismos que representan.

Posteriormente, estudio el costo de las evaluaciones para estimar cuánto se invirtió en esta práctica durante el periodo, y también para identificar en qué estrategias se canalizaron más recursos. Y por último, analizo la procedencia del financiamiento, uno de los elementos determinantes de la independencia del proceso evaluativo (*véase apartado 3.2*).

Obtuve la información que presento a través de una revisión documental de las evaluaciones, sus fichas técnicas y otras publicaciones de Coneval y Secretaría de Hacienda y Crédito Público (SHCP).

¿Cómo se determinó cuáles programas se evaluarían?

Los Lineamientos para la Evaluación publicados en 2007, establecen que los programas sujetos a reglas de operación (y otros que se determinen) deben ser evaluados y monitoreados, a fin de fomentar una gestión basada en resultados. Asimismo, señalan que el Coneval debe publicar, en cada ejercicio fiscal, un Programa Anual de Evaluación (PEA)²⁴ en colaboración con SHCP y Secretaría de la Función Pública (SFP) para dar a conocer los programas presupuestarios que serán valorados.

Por lo anterior, en términos formales, todas las evaluaciones realizadas a los programas de Sedesol fueron solicitadas por Coneval a través del PEA. Sin embargo, en la práctica, las dependencias y entidades también pueden solicitar de manera explícita la valoración de algún(os) programa(s). No existen registros públicos sobre estas solicitudes pero desde 2008 existe un inciso en el PEA que las posibilita.²⁵

Al indagar sobre las evaluaciones de diseño y desempeño realizadas a los programas de Sedesol, encontré que el Coneval da a conocer (en su portal de Internet) un total de 134 evaluaciones realizadas a 28 programas federales de Sedesol (véase *anexos*).²⁶ Del universo de evaluaciones disponible, analicé 111 que contaban con la valoración de GESOC (36 de diseño y 75 de desempeño) (véase *cuadro 2.1*).²⁷

Cuadro 2. 1 Evaluaciones de Sedesol con valoración de GESOC, 2007-2014

Evaluaciones	2007	2008	2009	2010	2011	2012	2013	2014	Total	%
Diseño	15				19		2		36	32
Específicas de Desempeño		12	18	16		13		16	75	68
Total	15	12	18	16	19	13	2	16	111	100

Fuente: Elaboración propia a partir de información de Coneval.

²⁴ En cumplimiento de los Lineamientos Generales y al Decreto de Presupuesto de Egresos de la Federación, se publica el Programa Anual de Evaluación (PEA) para dar a conocer las evaluaciones que se llevarán a cabo o comenzarán durante cada uno de los ejercicios fiscales y a los programas federales que aplican (CONEVAL, 2005 en: <http://www.coneval.gob.mx/Evaluacion/NME/Paginas/ProgramaAnual.aspx>).

²⁵ En la sección VII del PAE 2008 se especifica que las dependencias y entidades de la Administración Pública Federal responsables de operar programas federales, deberán enviar a la Secretaría, a la Función Pública y al Consejo: un listado de programas a su cargo que consideren deberán ser evaluados, así como las evaluaciones estratégicas a un programa o conjunto de programas que proponen para el siguiente ciclo presupuestario. Esta redacción fue simplificada a partir de 2009, señalando solo que el Consejo es responsable de atender las consultas y las solicitudes referentes al PAE en el ámbito de sus competencias

²⁶ Disponibles en la base de datos de CONEVAL sobre la evaluación de política social en: <http://www.coneval.gob.mx/Evaluacion/Paginas/InformeEvaluacion.aspx>

²⁷ Para algunos casos, GESOC señaló en sus informes que le fue imposible estimar los índices de algunos programas debido a problemas de opacidad en la información requerida.

Las 111 evaluaciones analizadas comprenden 23 programas de la Secretaría: 13 operados centralmente (Sedesol) y diez en colaboración con órganos administrativos desconcentrados (Diconsa, Liconsa, Fonart, Fonhapo, Indesol, Imjuventud, y la Coordinación Nacional del Programa de Desarrollo Humano Oportunidades) (véase cuadro 2.2).

Cuadro 2. 2 Programas federales de Sedesol considerado en el análisis de evaluaciones

No.	Nombre del Programa	Siglas	Tema	EED	ED	Total
Operados centralmente						
1	Programa Hábitat	Hábitat	Desarrollo	3	2	5
2	Programa para el Desarrollo Local (Microrregiones)	PDL	Desarrollo	1		1
3	Programa 3x1 para Migrantes	3X1	Desarrollo	5	2	7
4	Programa de Atención a Jornaleros Agrícolas	PAJA	Desarrollo	5	2	7
5	Programa Atención a Adultos Mayores en Zonas Rurales	PAAMZR	Desarrollo		1	1
6	Programa 70 y Más / Programa Pensión para Adultos Mayores	70 y Más	Desarrollo	5	3	8
7	Programa para el Desarrollo de Zonas Prioritarias	PDZP	Desarrollo	4	1	5
8	Programa Seguro de Vida para Jefas de Familia	PSVJF	Género	1	1	2
9	Programa de Opciones Productivas	POP	Productivo	4	2	6
10	Programa de Empleo Temporal	PET	Productivo	3	1	4
11	Programa de Estancias Infantiles para Apoyar a Madres Trabajadoras	PEI	Productivo	4	2	6
12	Programa de Rescate de Espacios Públicos	PREP	Vivienda o Espacios Públicos	1	2	3
13	Programa de Apoyo a los AVECINDADOS en Condiciones de Pobreza Patrimonial para Regularizar Asentamientos Humanos Irregulares	PASPRAH	Vivienda o Espacios Públicos	2	1	3
Operados por órganos desconcentrados						
14	Subsidio a Programas para Jóvenes	PSPJ	Desarrollo	1		1
15	Programa de Abasto Social de Leche a cargo de Liconsa, S.A. de C.V.	PASL	Desarrollo	5	2	7
16	Programa de Abasto Rural a cargo de Diconsa, S.A. de C.V.	PAR	Desarrollo	4	2	6
17	Programa de Desarrollo Humano Oportunidades / PROSPERA	PROSPERA	Desarrollo	4	1	5
18	Programa de Apoyo Alimentario	PAL	Desarrollo	5	2	7
19	Programa de Apoyo a las Instancias de Mujeres en las Entidades Federativas para Implementar y Ejecutar Programas de Prevención de la Violencia Contra las Mujeres	PAIMEF	Género	4	1	5
20	Programas del Fondo Nacional de Fomento a las Artesanías	FONART	Productivo	5	2	7
21	Programa de Coinversión Social	PCS	Productivo	4	2	6
22	Programa de Ahorro, Subsidio y Crédito para la Vivienda "Tu Casa"	PASCV	Vivienda o Espacios Públicos	3	2	5
23	Programa de Vivienda Rural	PVR	Vivienda o Espacios Públicos	2	2	4
				75	36	111

Fuente: Elaboración propia con datos de Coneval.

Nota: EED – Evaluación Específica de Desempeño, ED – Evaluación de Diseño.

En cuanto al tipo de evaluaciones aplicadas, en apego a los Lineamientos y al cronograma de los PEAs, las Evaluaciones de Diseño (ED) se elaboraron durante el primer año de operación de los programas y cuando se solicitaron las evaluaciones de consistencia y resultados (cuando inician las evaluaciones y en el cambio de sexenio, 2007 y 2011 respectivamente). Por otro lado, las Evaluaciones Específicas de Desempeño (EED) se realizaron anualmente a partir del 2008, pero sin coincidir con las ED.

El programa que sumó más evaluaciones fue 70 y Más, con ocho valoraciones (cinco EED y tres ED); seguido de 3x1, PAJA, PASL, PAL, y FONART, con siete valoraciones cada uno en total.

Para enriquecer este análisis clasifiqué los 23 programas en cuatro categorías, a partir del tema que atienden: desarrollo (52 por ciento), productivos (21.7 por ciento), vivienda o espacios públicos (17 por ciento), y género (8.7 por ciento).²⁸

La clasificación que sugiero me permitió analizar las evaluaciones asumiendo que existen diferentes retos conceptuales y operativos en los programas sociales, dependiendo del tema o problema público al que se abocan.

²⁸ La clasificación de los programas federales de Sedesol, según su orientación al desarrollo y a fines productivos, se retomó de una distinción que hace Inmujeres en una de sus páginas de internet. Publicado en: <http://padrones.inmujeres.gob.mx/listaProgramas.php>. Adicionalmente, incorporé las dos categorías (vivienda o espacios públicos, y género) a esta investigación.

2.1.1. ORGANISMOS Y EXPERTOS EVALUADORES

En esta sección analizo qué organismos evaluadores y expertos en la materia llevaron a cabo las valoraciones a los programas de Sedesol. Coneval es el responsable de coordinar las evaluaciones de Sedesol. No obstante, dentro de la Secretaría existe una unidad administrativa independiente, también conocida como unidad de evaluación, que se encarga de contratar y supervisar la realización de las valoraciones.²⁹

La regulación vigente en México establece que dependencias y entidades deben evaluar los programas públicos por conducto de organismos evaluadores independientes, pudiendo participar: instituciones académicas y de investigación, personas físicas o morales especializadas en la materia u organismos de carácter nacional o internacional que cuenten con reconocimiento y experiencia, en términos de las disposiciones aplicables (Art. 73, LGDS, 2004; Art. 78, LFPRH, 2006; Norma difusión, 2013).³⁰

En el análisis efectuado identifiqué la participación de 27 organismos evaluadores (*véase anexos*) y los agrupé en nueve categorías según su naturaleza: colegios, centros de estudio, centros de investigación, universidades, institutos nacionales, sociedad civil, consultores, consultores independientes u organismos internacionales. En el *gráfico 2.1* se puede observar que colegios,³¹ centros de estudio³² y centros de investigación³³ concentraron casi 58 por ciento de las evaluaciones.

Por otra parte, en el *gráfico 2.2* se aprecia que de las 111 evaluaciones analizadas el cien por ciento de los contratos otorgados a consultores independientes y centros de estudio fueron para EED; mientras que a los consultores privados solo se les consideró para ED. El resto de los organismos evaluadores realizaron ambos tipos de valoraciones.

²⁹ Generalmente las dependencias y entidades valoradas financian las evaluaciones. Sin embargo, cuando lo hace CONEVAL, éste desempeña una participación activa en la contratación y supervisión de dichas evaluaciones.

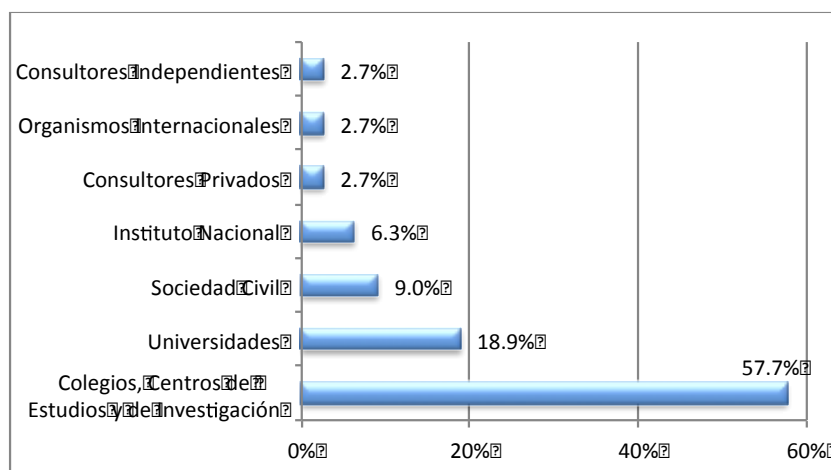
³⁰ Para pactar la participación de estos organismos evaluadores, las dependencias y entidades deben, sin modificar su estructura, designar una unidad administrativa ajena a la operación de los programas para coordinar la contratación, operación y supervisión de las evaluaciones (sección 29, RLFPRH, 2006).

³¹ COLMEX (30), Colegio Mexiquense (8), COLEF (5), Colegio de Economistas (1), y Colegio de Postgraduados(1).

³² CEESS (11)

³³ CIDE (4), CIESAS (3) y CIAD (1).

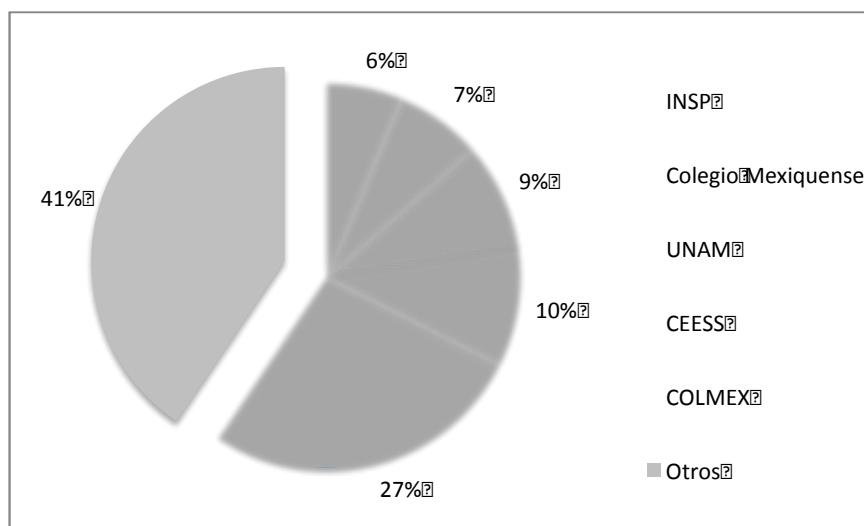
Gráfico 2. 1 Porcentaje de evaluaciones realizadas por tipo de organismo evaluador



Fuente: Elaboración propia con datos de Coneval.

Llevando a cabo un análisis más detallado, hallé que solo en cinco organismos evaluadores (COLMEX, CEESS, UNAM, Colegio Mexiquense, y INSP) se concentra casi 60 por ciento de las evaluaciones analizadas. Destacando la participación del COLMEX³⁴ como el evaluador más recurrente, puesto que recibió el encargo de casi una tercera parte de dichas valoraciones entre 2007 y 2014.

Gráfico 2. 2 Organismos evaluadores con mayor participación en la valoración de los programas federales de Sedesol



Fuente: Elaboración propia con datos de Coneval.

³⁴ El Colegio de México es una institución pública, de carácter universitario, dedicada a la investigación y a la enseñanza superior.

Además, el COLMEX fue el único organismo que realizó más de 50 por ciento de las valoraciones en cinco programas. Destacando el caso del Programa para el Desarrollo de Zonas Prioritarias (PDZP), para el cual realizó 100 por ciento de las evaluaciones (véase cuadro 2.3).

Después de conocer qué organismos evaluaron los programas de Sedesol, estimé necesario identificar a los expertos que coordinaron dichas evaluaciones para tener una idea de qué tan amplia es la oferta de personal capacitado dentro del sector. Esto debido a que, principalmente, la práctica de evaluación en México es relativamente nueva y que los espacios para la profesionalización en esta materia, a nivel nacional, están apenas en desarrollo.

Cuadro 2. 3 Casos en que un mismo evaluador realiza más de 50 por ciento de las valoraciones de un programa presupuestario

No.	Nombre del Programa	Evaluador*	Total Evaluaciones	Evaluaciones revisadas	Porcentaje
1	Programa de Apoyo a los Avecindados en Condiciones de Pobreza Patrimonial para Regularizar Asentamientos Humanos Irregulares	CEDUA, COLMEX ^a	3	2	66.7
2	Programa 3x1 para Migrantes	CEE y CEDUA, COLMEX ^b	7	5	71.4
3	Programa de Vivienda Rural	CEE y CEDUA, COLMEX ^c	4	3	75.0
4	Programa de Ahorro, Subsidio y Crédito para la Vivienda "Tu Casa"	CEE y CEDUA, COLMEX ^d	5	4	80.0
5	Programa para el Desarrollo de Zonas Prioritarias	CEE y CEDUA, COLMEX ^e	5	5	100.0

Fuente: Elaboración propia con datos de Coneval.

* **Nota:** CEE - Centro de Estudios Económicos; CEDUA – Centro de Estudios Demográficos, Urbanos y Ambientales.

a) Martha Rosa Schteingart Garfunkel (2009) y Vicente Ugalde Saldaña (2010).

b) Gerardo Esquivel Hernández (2011), Martha Rosa Schteingart Garfunkel (2009, 2010 y 2012) y Vicente Ugalde Saldaña (2014).

c) Gerardo Esquivel Hernández (2007), Martha Rosa Schteingart Garfunkel (2009) y Vicente Ugalde Saldaña (2010).

d) Gerardo Esquivel Hernández (2008 y 2011), Martha Rosa Schteingart Garfunkel (2009) y Vicente Ugalde Saldaña (2010).

e) Gerardo Esquivel Hernández (2011), Martha Rosa Schteingart Garfunkel (2009 y 2010), Clara Eugenia Salazar Cruz (2012) y Luis Jaime Sobrino Figueroa (2014).

En este sentido, al orientar el análisis hacia los expertos responsables de las evaluaciones, identifiqué la participación de 40 coordinadores (véase anexos). Además, observé que cinco de ellos colaboraron con más de un organismo evaluador (véase cuadro 2.4). El caso más representativo fue el de Israel Banegas González, quien

coordinó el mayor número de evaluaciones entre 2007 y 2014, a través de dos organismos (Colegio Mexiquense y UNAM).

Cuadro 2. 4 Coordinadores de evaluaciones que participaron con más de una institución para valorar los programas de Sedesol

No.	Nombre del Coordinador	Institución evaluadora	Número evaluaciones	Total evaluaciones
1	Carola Conde Bonfil	UNAM	1	2
		Colegio Mexiquense	1	
2	Fernando Butler Silva	Colegio de Economistas	1	2
		UAEM	1	
3	María Guadalupe Alcocer Villanueva	Alternativas en Economía	1	2
		CEPNA	1	
4	María Begoña Iguñiz Echeverría	BETAKOROSI CONSULTORES	1	3
		NIK BETA	2	
5	Israel Banegas González	Colegio Mexiquense	7	17
		UNAM	10	

Fuente: Elaboración propia con datos de Coneval.

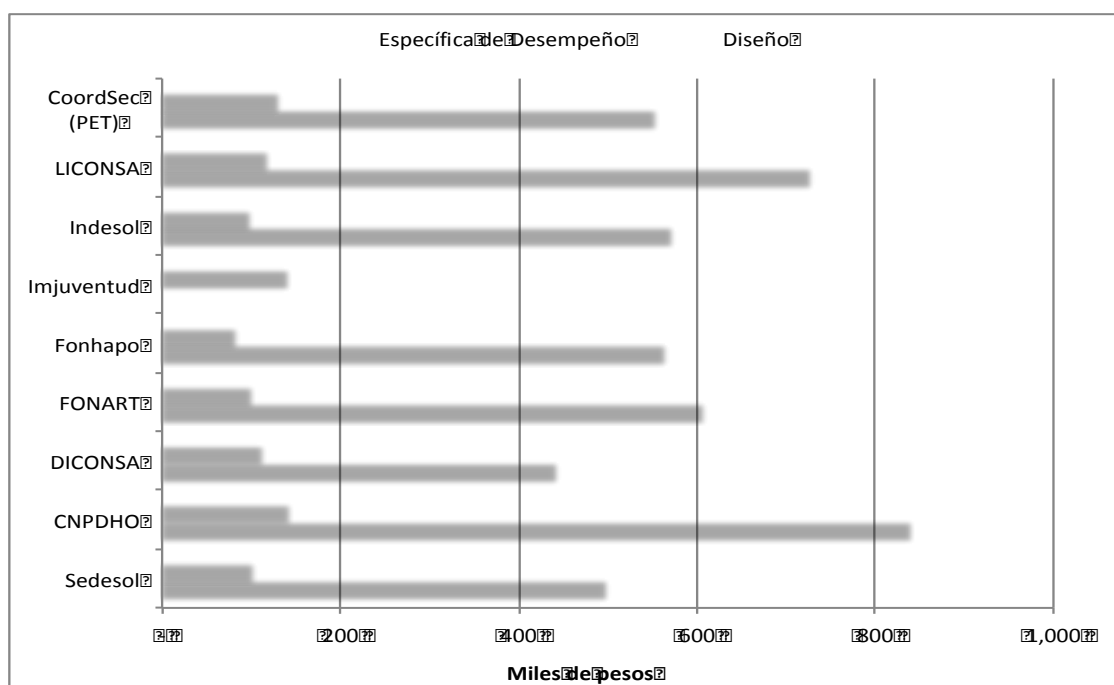
Por último, cabe agregar que más de 35 por ciento de las evaluaciones analizadas estuvieron a cargo de solo tres coordinadores. Como ya se mencionó, Israel Banegas fue el que más participó, coordinando 15 por ciento; seguido de Martha R. Schteingart Garfunkel y Evelyne E. Rodríguez Ortega, quienes coordinaron aproximadamente diez por ciento de las evaluaciones cada una (*véase anexos*).

2.1.2. EL COSTO

En esta sección, analizo el costo de las evaluaciones. Estimo cuánto se invirtió durante el periodo, y también identifico en qué estrategias se canalizaron más recursos. Después de una revisión documental de los informes de evaluación y de las bases de datos de Coneval, encontré que el costo de las 111 evaluaciones que analizo en esta tesis, fue de 26.99 millones de pesos: 71 por ciento en las ED (19.15 millones) y 29 en las EED (7.84 millones).

Tomando en cuenta que más de 70 por ciento del costo estimado se destinó a las ED, que representa solo una tercera parte de las evaluaciones analizadas (*véase cuadro 2.1*), resulta evidente que las ED fueron significativamente más onerosas que las EED, esto podría deberse a que las primeras son más exhaustivas. En el *gráfico 2.4* se muestra la diferencia en el costo promedio según el tipo de evaluación, pero también cómo dicho costo varía de acuerdo a la procedencia del financiamiento.³⁵

Gráfico 2. 3 Costo promedio de las evaluaciones realizadas según tipo de evaluación y procedencia del financiamiento, 2007 - 2014



Fuente: Elaboración propia con datos de Coneval.

³⁵ El financiamiento para las evaluaciones puede provenir de Sedesol, los órganos desconcentrados de esta secretaría o de CONEVAL, cuando este último así lo estipule.

Las evaluaciones de diseño financiadas por la Coordinación Nacional del Programa de Desarrollo Humano Oportunidades (CNPDHO) y por LICONSA fueron las que registraron el costo promedio más alto. Sin embargo, al margen de estos dos casos, la variación que se observa entre dependencias en el costo por tipo de evaluación, no es muy significativa (véase gráfico 2.4).

Por otra parte, examinando el costo total de las evaluaciones, identifiqué que sí existe una gran diferencia en los montos gastados tanto por tipo de evaluación como por fuente de financiamiento. En el *cuadro 2.5*, se puede apreciar que las ED fueron las que tuvieron mayor variación en su costo, oscilando entre 200 mil y un millón de pesos. El monto más bajo lo registraron dos programas operados por Sedesol (Seguro de Vida para Jefas de Familia (PSVJF) y 70 y Más en 2013),³⁶ mientras que el monto más alto fue por el Programa para el Desarrollo Humano Oportunidades (PDHO) a cargo de la CNPDHO durante 2011.³⁷

Cuadro 2. 5 Costo total de las evaluaciones de Sedesol según tipo de evaluación, 2007-2014

	Diseño (costo en miles de pesos)				Específica de Desempeño (costo en miles de pesos)			
	Total	Máximo	Medio	Mínimo	Total	Máximo	Medio	Mínimo
Sedesol	8,963.4	783.6	498.0	208.8	4,044.9	182.0	101.1	70.0
CNPDHO	1,678.0	1,044.0	839.0	634.0	568.3	182.0	142.1	87.6
Diconsa	1,324.8	507.8	441.6	398.3	672.7	182.0	112.1	83.0
FONART	1,212.9	638.0	606.4	574.9	497.4	135.0	99.5	74.1
Fonhapo	2,252.2	783.0	563.1	395.0	409.5	97.0	81.9	74.1
Imjuventud	-	-	-	-	140.0	140.0	140.0	140.0
Indesol	1,713.0	638.0	571.0	500.0	786.1	140.0	98.3	70.0
LICONSA	1,452.7	829.4	726.4	623.3	589.7	182.0	117.9	83.0
CoordSec (PET)	552.9	552.9	552.9	552.9	129.9	129.9	129.9	129.9
Total	19,149.9	1,044.0	547.1	208.8	7,838.6	182.0	104.5	70.0

Fuente: Elaboración propia con datos de Coneval.

En las EEDs la variación en el costo fue menor, osciló entre 70 y 182 mil pesos. El monto más bajo lo registraron cuatro programas durante 2010: Atención a Jornaleros

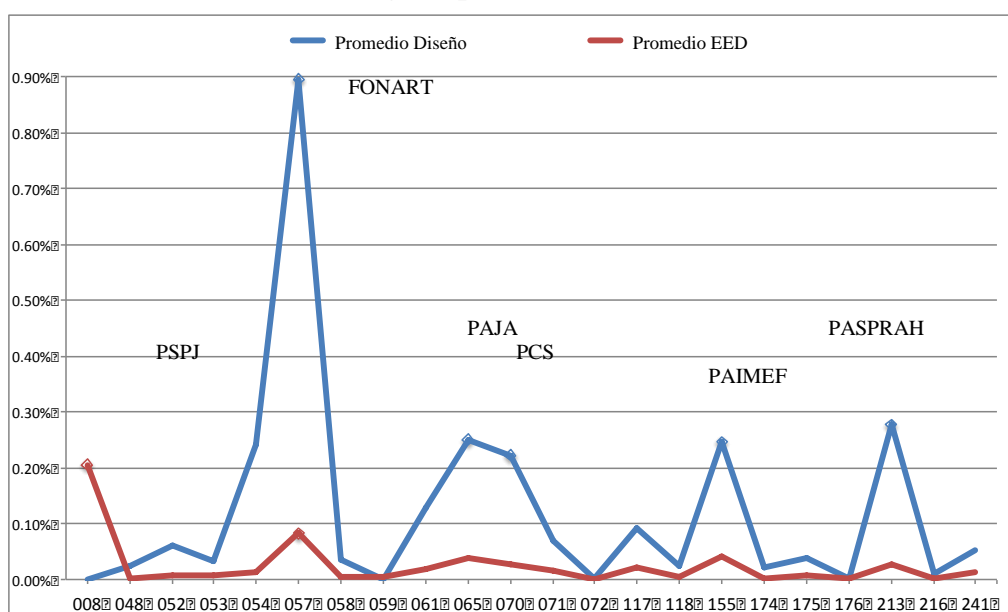
³⁶ Dos programas valorados por Evalware Expertos en Evaluación de Políticas Públicas S.A. de C.V., y coordinados por Julio Franco Corzo y Torres.

³⁷ Evaluación realizada por el Colegio de México (COLMEX), y coordinada por Gerardo Esquivel Hernández.

Agrícolas (PAJA), Coinversión Social (PCS), Programa para la Prevención de la Violencia contra las Mujeres (PAIMEF), y 70 y Más.³⁸ El monto más alto lo obtuvieron otros cuatro programas en 2014: Abasto Social de Leche de LICONSA, Abasto Rural de Diconsa, Apoyo Alimentario y PDHO, los dos últimos a cargo de la CNPDHO.³⁹

Al estudiar la relación entre el costo de las evaluaciones y el presupuesto anual que de cada programa noté que, en la mayoría de los 111 casos analizados, el costo representa una proporción no mayor al 0.05 por ciento (*véase gráfico 2.5*). Cabe aclarar que durante la investigación no encontré información que me ayudara a explicar por qué la proporción del presupuesto que se asignó para las realización de las evaluaciones variaba y que en la legislación vigente tampoco se estipulan criterios para la asignación de dicho presupuesto.

Gráfico 2. 4 Proporción del presupuesto de los programas que representa el costo de las evaluaciones según tipo de evaluación, 2007 – 2014



Fuente: Elaboración propia con datos de Coneval.

³⁸ Cuatro programas que fueron evaluados por C-230 Consultores, y coordinado por Raúl Abreu Lastra.

³⁹ Cuatro programas evaluados por el Instituto Nacional de Salud Pública (INSP) y coordinados por Juan Pablo Gutiérrez Reyes.

2.1.3. LA FUENTE DE FINANCIAMIENTO

La regulación vigente establece que el costo de la evaluación deberá cubrirse con cargo al presupuesto de la dependencia o entidad correspondiente (Art. 180, RLFPRH, 2006). No obstante, el PAE señala que cuando Coneval es quien realiza la evaluación, el pago también puede ser con cargo a este organismo, cuando así se determine.

En el *cuadro 2.6*, a partir del contenido de los PAEs, muestro quién financió las evaluaciones durante el periodo analizado. Como se puede observar, todas las ED se cubrieron con recursos de Sedesol y sus órganos desconcentrados. En el caso de las EED, el Coneval solo financió los primeros tres años, dejando el pago de los últimos dos también a Sedesol y sus desconcentrados.

Cuadro 2.6 Origen del presupuesto de las evaluaciones por año, según tipo de evaluación

	2007	2008	2009	2010	2011	2012	2013	2014
Evaluación de Diseño (Consistencia y Resultados)	DyE				DyE			
Evaluación de Diseño (Programas nuevos)	DyE				DyE		DyE	
Específicas de Desempeño		Coneval	Coneval	Coneval		DyE		DyE

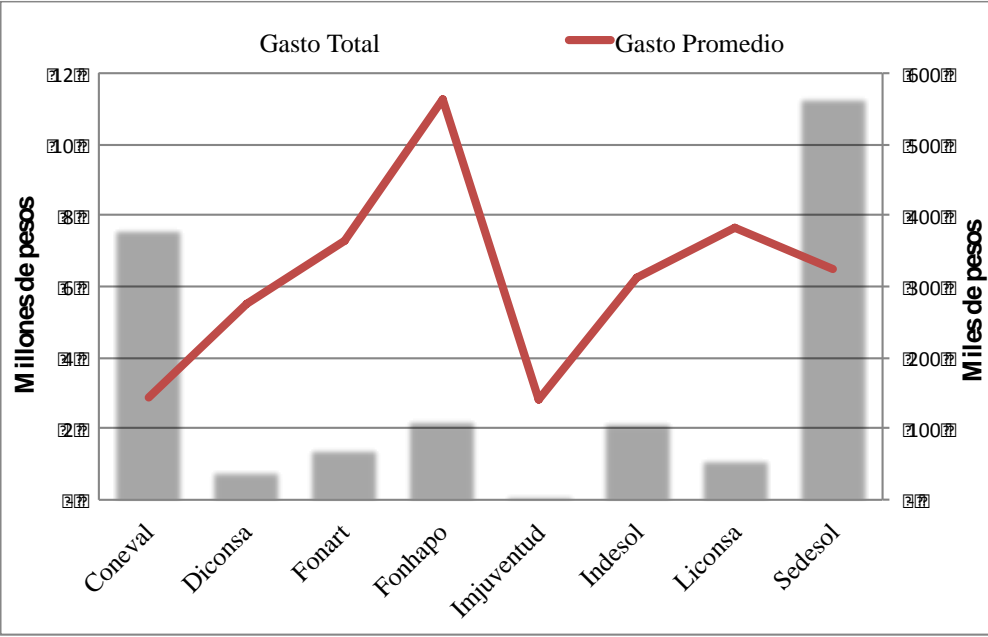
Nota: “DyE”, financiamiento a cargo de las Dependencias y Entidades que operan los programas evaluados.

Fuente: Elaboración propia con datos del Programa Anual de Evaluación, 2007-2014.

Considero relevante mencionar que, si bien todas las evaluaciones realizadas a los programas de Sedesol son coordinadas por Coneval, solo cuando éste último asume los costos es cuando desempeña un papel activo en la contratación y supervisión del proceso evaluativo. En el resto de los casos, solo funge como una especie de gestor que verifica el cumplimiento de la normatividad aplicable.

Al explorar el origen del presupuesto que se destinó a las evaluaciones, noté que Sedesol y Coneval fueron las que aportaron más recursos: once y siete millones de pesos respectivamente. También fueron las que financiaron más evaluaciones: Sedesol 32 por ciento, y Coneval 53 por ciento. Por su parte, Fonhapo fue la que registró el gasto promedio más alto (*véase gráfico 2.6*).

Gráfico 2. 5 Gasto total y promedio de las evaluaciones según procedencia del presupuesto



Fuente: Elaboración propia con datos del Programa Anual de Evaluación.

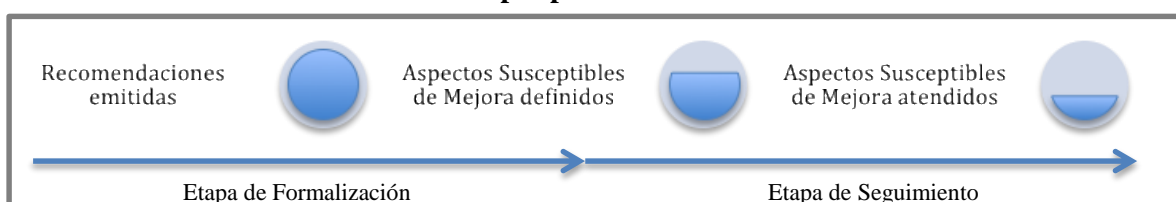
2.2. LA UTILIZACIÓN DE LAS EVALUACIONES

En este apartado, presento un análisis descriptivo sobre la utilización de las evaluaciones con el objetivo de documentar en qué medida las recomendaciones emitidas fueron empleadas para mejorar los programas de Sedesol a lo largo de los años.

En México, al concluir el proceso evaluativo, las recomendaciones emitidas son revisadas por la dependencia o entidad valorada y, a partir de esta revisión, la unidad responsable del programa pacta un compromiso para atender aquellas que determina pertinentes. Posteriormente, se inicia un proceso de seguimiento del que se desprenden informes anuales sobre el avance en su consecución.

Retomando estos pasos que siguen los programa para atender las recomendaciones, estudié la utilización a través de dos etapas: en la primera, calculo la proporción de recomendaciones emitidas que logró ser formalizada y se volvió obligatoria⁴⁰ y, en la segunda, identifiqué la proporción de recomendaciones formalizadas que efectivamente fue atendida (véase *ilustración 2.2*).

Ilustración 2. 2 Diferentes etapas para la utilización de las evaluaciones



Fuente: Elaboración propia.

En la regulación mexicana están implícitas las dos etapas que acabo de señalar. Sin embargo, la utilización de las evaluaciones, se monitorea solo en la segunda etapa, a través de informes anuales, debido a que esta permite observar efectos directos en los programas. Es decir que usualmente se considera este aspecto para determinar la efectividad de una evaluación.

⁴⁰ En México, las recomendaciones formalizadas cuyo cumplimiento se vuelve obligatorio son denominadas Aspectos Susceptibles de Mejora (ASM).

No obstante, a pesar de que la fase de formalización no ayuda a explicar el efecto de una evaluación, consideré pertinente incorporar la proporción de recomendaciones que no fueron utilizadas. En otras palabras, que se excluyen del proceso de retroalimentación para el cual fueron emitidas, con la finalidad de ayudar a mejorar los programas.⁴¹

Los datos que empleé en el análisis (número de recomendaciones emitidas, formalizadas y atendidas) se obtuvieron de dos fuentes: por un lado, un análisis cualitativo original que llevé a cabo para las 111 evaluaciones y, por otro, los informes anuales publicados por Sedesol donde se reportan los avances en el cumplimiento de las recomendaciones formalizadas. A continuación menciono de manera concisa las especificaciones regulatorias señaladas para cada etapa y los resultados obtenidos.

⁴¹ Cabe aclarar que no se da por hecho que las recomendaciones que no fueron utilizadas con fines instrumentales, para el mejoramiento del programa valorado, no hayan tenido su contribución en el ámbito conceptual al proporcionar un mejor entendimiento sobre las dificultades y los retos del programa en cuestión.

2.2.1. DE RECOMENDACIONES A ASPECTOS SUSCEPTIBLES DE MEJORA

En México, los Lineamientos 2007 (SFP, 2015) establecen que una vez concluida la evaluación, las dependencias y entidades deben considerar los resultados obtenidos para seleccionar y clasificar sus Aspectos Susceptibles de Mejora (ASM). En este sentido, los ASM son definidos como las acciones con las cuales se dará atención a los “hallazgos, debilidades, oportunidades y amenazas identificadas en las evaluaciones externas y/o informes, que pueden ser atendidos por el programa” (Mecanismo de seguimiento, 2011).

En Sedesol, la Dirección General de Evaluación y Monitoreo de Programas Sociales (DGEMPS) se encarga de supervisar la elaboración de las evaluaciones que se realizan en la dependencia, así como de gestionar la formalización de ASM en coordinación con las unidades responsables de los programas valorados. Cabe aclarar que éstos últimos no están obligados a formalizar ASM ni a seguir criterios específicos de selección. Las únicas características que deben cumplir los ASM son: claridad, relevancia, justificación y factibilidad.

En general, durante el periodo analizado, el porcentaje de recomendaciones que se formalizó como ASM disminuyó sostenidamente a lo largo del tiempo, pasando de 70 por ciento en 2007 a 58 en 2014. A continuación, se analiza esta tendencia tomando en consideración: el tipo de evaluación (diseño o desempeño), la temática de los programas (desarrollo, vivienda o espacios públicos, género o productivos), y la formalización por cada programa.

(a) LA FORMALIZACIÓN SEGÚN TIPO DE EVALUACIÓN

La disminución observada en la proporción de recomendaciones formalizadas fue mayor en las ED, pasando de 71 a 50 por ciento del 2007 al 2011.⁴² Esta tendencia también se observa en las EED aunque en menor proporción, descendiendo de 66 a 58

⁴² En este comparativo se excluyen los resultados del año 2013 debido a que solo se realizan evaluaciones de diseño a programas de nueva creación y el análisis solo incluye dos.

por ciento. Las EED alcanzaron su nivel más bajo en 2010 con 45 por ciento (*véase gráfico 2.7*).

En el caso de las ED, si bien los programas presentaron un declive generalizado para el 2011, este cambio se explica en gran medida por los bajos porcentajes reportados en los siguientes programas: Vivienda Rural (9 por ciento), Atención a Jornaleros Agrícolas (14 por ciento), y Opciones Productivas (16.6 por ciento).

Por otro lado, la disminución que registraron las EED en 2010 se puede explicar por una abrupta caída en dos programas: Apoyo Alimentario (0 por ciento), y Fondo Nacional de Fomento a las Artesanías (14 por ciento).

(b) LA FORMALIZACIÓN POR TEMÁTICA

Al analizar la proporción de recomendaciones formalizadas a partir de los cuatro temas centrales que abordan los programas (desarrollo, vivienda o espacios públicos, productividad, y género), observé que aquellos orientados al desarrollo y a la vivienda o espacios públicos fueron los que registraron los porcentajes más altos, ambos con 61 por ciento.

En un análisis más detallado, pude apreciar que los programas orientados al desarrollo atendieron en mayor medida las recomendaciones derivadas de las ED, mientras que en los programas dirigidos a la vivienda o los espacios públicos se formalizaron más las provenientes de EED (*véase gráfico 2.9*). De la tendencia observada podría asumir que los programas de desarrollo tienden a formalizar las recomendaciones de diseño porque demandan una mayor complejidad conceptual.

En el *gráfico 2.11*, se aprecia que la disminución en el porcentaje de recomendaciones formalizadas se enfatizó en los programas de género durante 2011, y en los de desarrollo en 2013. En ambos casos se trata de recomendaciones de diseño.

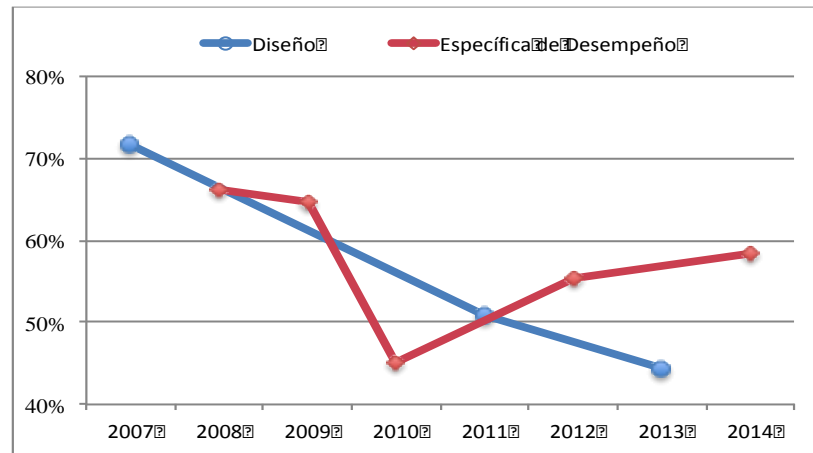
La caída registrada en los programas de género (2011) y en los de desarrollo (2013) se debe, en ambos casos, al valor de un solo programa que reportó un porcentaje muy bajo: en género, el Programa de Prevención de la Violencia contra las Mujeres⁴³ que descendió ese año de 50 a 28.6 por ciento, y en desarrollo, el Programa 70 y Más, que descendió de 43 a 22 por ciento.

(c) LA FORMALIZACIÓN POR PROGRAMA

En promedio, entre 2007 y 2014, se volvieron ASM 57.9 por ciento de las recomendaciones a los programas de Sedesol. Los dos programas que reportaron un mayor porcentaje fueron: 3x1 para Migrantes (86.4 por ciento), y Ahorro, Subsidio y Crédito para la Vivienda “Tu Casa” (86.3 por ciento). Por otra parte, el programa que registró el menor porcentaje de recomendaciones formalizadas fue el de Guarderías y Estancias Infantiles para Apoyar a Madres Trabajadoras (38.5 por ciento), seguido de Opciones Productivas (41.9 por ciento) (*véase gráfico 2.13*).

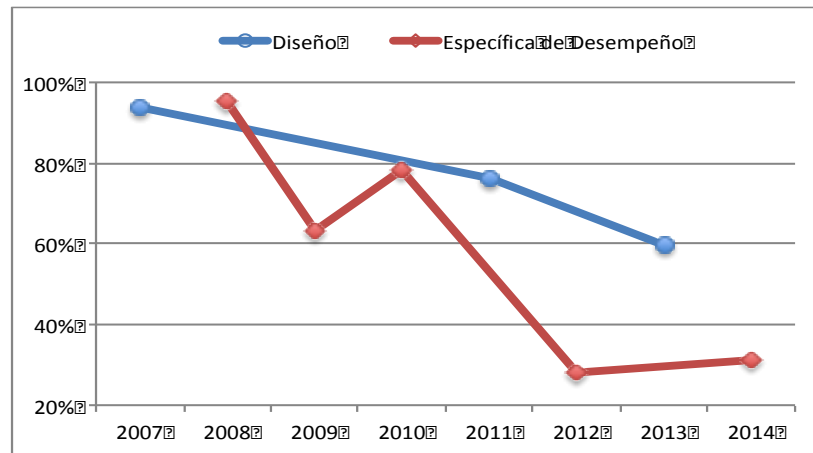
⁴³ El nombre completo es Programa de Apoyo a las Instancias de Mujeres en las Entidades Federativas para Implementar y Ejecutar Programas de Prevención de la Violencia contra las Mujeres.

Gráfico 2. 6 Porcentaje de recomendaciones que se convierten en ASM por tipo de evaluación



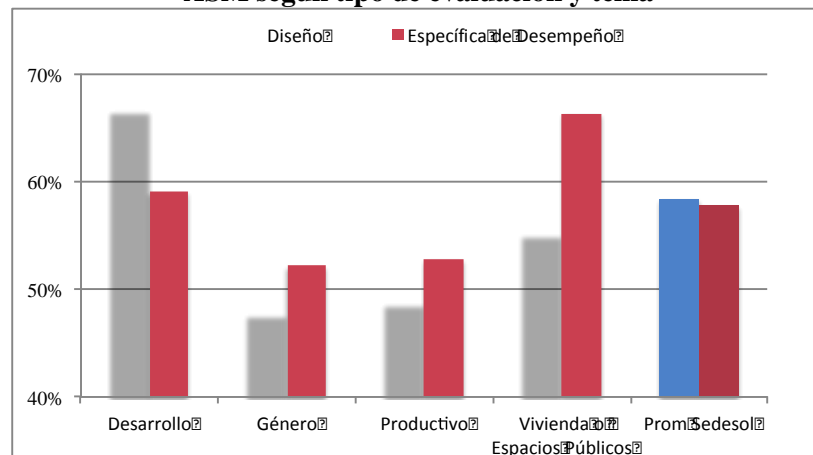
Fuente: Elaboración propia con datos de las evaluaciones y los informes de avances.

Gráfico 2. 7 Porcentaje de ASM atendidos por tipo de evaluación



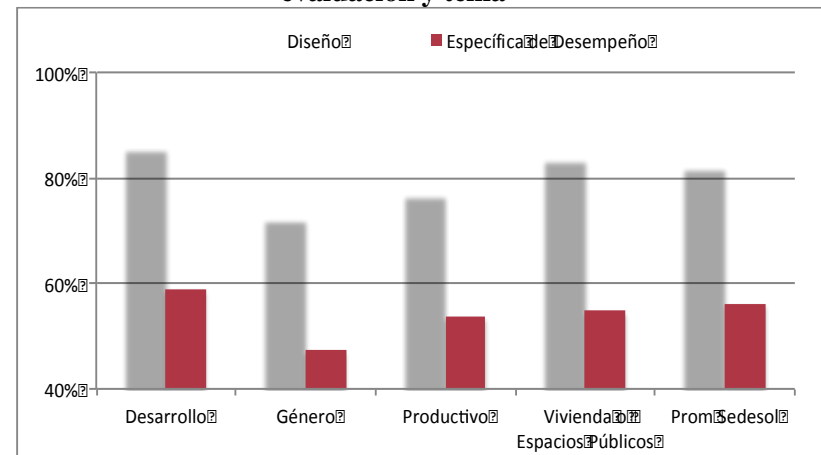
Fuente: Elaboración propia con datos de las evaluaciones y los informes de avances.

Gráfico 2. 8 Porcentaje de recomendaciones que se convierten en ASM según tipo de evaluación y tema



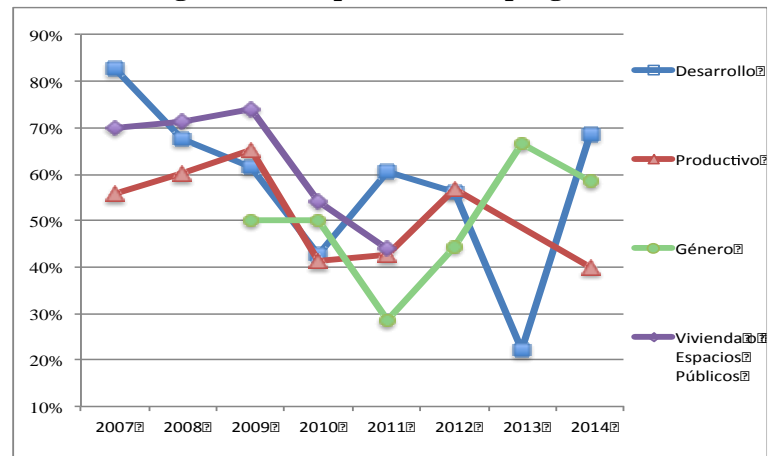
Fuente: Elaboración propia con datos de las evaluaciones y los informes de avances.

Gráfico 2. 9 Porcentaje de ASM atendidos según tipo de evaluación y tema



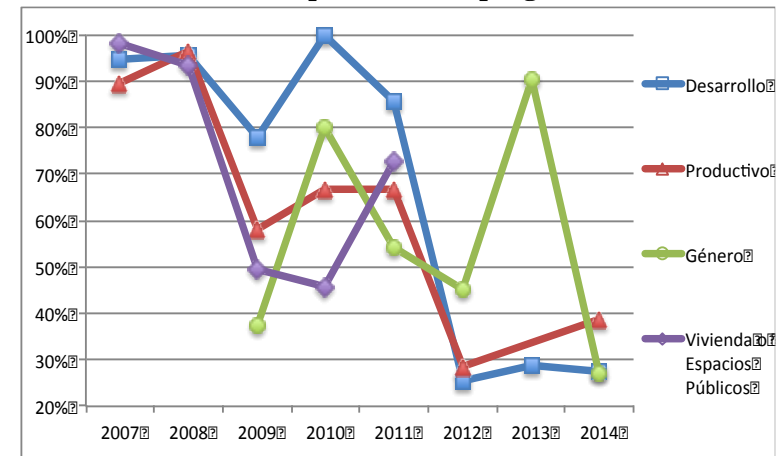
Fuente: Elaboración propia con datos de las evaluaciones y los informes de avances.

Gráfico 2. 10 Porcentaje de recomendaciones que se convierten en ASM según el tema que aborda el programa



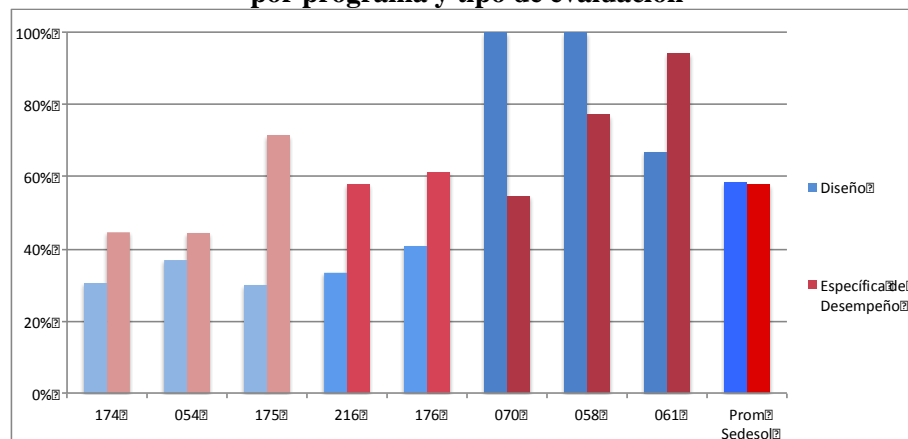
Fuente: Elaboración propia con datos de las evaluaciones y los informes de avances.

Gráfico 2. 11 Porcentaje de ASM atendidos según el tema que aborda el programa



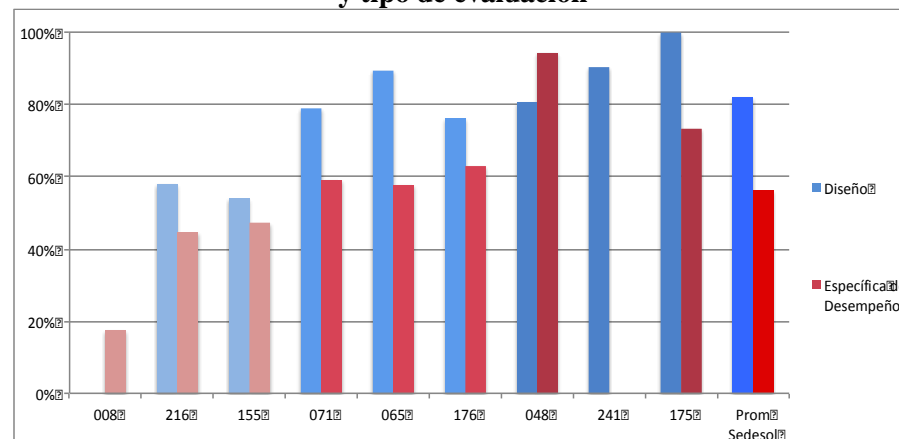
Fuente: Elaboración propia con datos de las evaluaciones y los informes de avances.

Gráfico 2. 12 Porcentaje de recomendaciones que se convierten en ASM por programa y tipo de evaluación



Fuente: Elaboración propia con datos de las evaluaciones y los informes de avances.

Gráfico 2. 13 Porcentaje de ASM atendidos por programa y tipo de evaluación



Fuente: Elaboración propia con datos de las evaluaciones y los informes de avances.

Porcentaje de recomendaciones que se convierten en ASM			
174	Programa de Guarderías y Estancias Infantiles para Apoyar a Madres Trabajadoras	Productivo	38.5%
054	Programa de Opciones Productivas	Productivo	41.9%
175	Programa de Rescate de Espacios Públicos	Vivienda o Espacio Púb	43.7%
216	Programa para el Desarrollo de Zonas Prioritarias	Desarrollo	53.0%
176	Programa 70 y Más	Desarrollo	53.6%
070	Programa de Coinversión Social	Productivo	69.7%
058	Programa de Ahorro, Subsidio y Crédito para la Vivienda "Tu Casa"	Vivienda Espacio Púb	86.3%
061	Programa 3x1 para Migrantes	Desarrollo	86.4%

Porcentaje de ASM atendidos			
008	Subsidio a Programas para Jóvenes	Desarrollo	17.5%
216	Programa para el Desarrollo de Zonas Prioritarias	Desarrollo	47.3%
155	Programa de Apoyo a las Instancias de Mujeres en las Entidades Federativas para Prevención de la Violencia	Género	48.7%
071	Programa de Empleo Temporal	Productivo	64.1%
065	Programa de Atención a Jornaleros Agrícolas	Desarrollo	66.9%
176	Programa 70 y Más	Desarrollo	68.0%
048	Programa Hábitat	Desarrollo	87.5%
241	Programa Seguro de Vida para Jefas de Familia	Género	90.4%
175	Programa de Rescate de Espacios Públicos	Vivienda o Espacios Púb	91.1%

2.2.2. LA ATENCIÓN DE LOS ASPECTOS SUSCEPTIBLES DE MEJORA

En México, una vez formalizadas las recomendaciones como ASM, las dependencias y entidades valoradas se comprometen a darles seguimiento a través de la celebración de un convenio para el mejoramiento de la gestión para resultados. Su cumplimiento es supervisado por SFP (Lineamientos, 2007; Función Pública, 2015).

Para la adecuada atención de los ASM, las dependencias y entidades deben cumplir con el Mecanismo para el Seguimiento a los Aspectos Susceptibles de Mejora (Mecanismo de Seguimiento) emitido por Coneval, en conjunto con SHCP y SFP. Este Mecanismo de Seguimiento busca hacer efectiva la utilización de la evaluación al definir responsables, formalizar instrumentos de trabajo y fechas de entrega de los documentos (Coneval, 2005).⁴⁴

En términos generales, el Mecanismo de Seguimiento establece el procedimiento que se debe observar para suscribir y supervisar los ASM, con el fin de mejorar la política y los programas sociales. Además, define cuatro tipos de ASM de acuerdo a qué actores se encuentran involucrados en su solución y el nivel de prioridad: específicos (corresponden a unidades responsables); institucionales (una o más áreas de la dependencia); interinstitucionales (más de una dependencia o entidad); e intergubernamentales (demandan la intervención de gobiernos estatales o municipales).

En este apartado sólo considero los ASM específicos e institucionales que fueron atendidos y reportados por Sedesol en sus informes de avance, entre 2007 y 2014.

(a) LA ATENCIÓN DE ASM SEGÚN TIPO DE EVALUACIÓN

Al analizar el cumplimiento de los ASM, observé que las ED registraron un porcentaje promedio mayor. En este sentido, mientras que las ED atendieron 82 por ciento de los ASM, las EED alcanzaron apenas 56 por ciento (*véase gráfico 2.8*).

⁴⁴ Derivado de este proceso, el CONEVAL presenta un Informe de Seguimiento que incluye los principales hallazgos del proceso así como un conjunto de recomendaciones que requieren la atención de la Comisión Intersectorial de Desarrollo Social y de la Comisión Nacional de Desarrollo Social (CONEVAL, 2005).

Al incumplirse más los ASM de las EED, asumo que esto se debe a una mayor complejidad en las recomendaciones orientadas a la operación de los programas, respecto a las de diseño. Sin embargo, en una entrevista, el funcionario de Coneval “A” comentó que en términos prácticos no se espera que las EED ayuden a mejorar directamente los programas sino a proveer información a funcionarios de alto nivel. Un enfoque que pudiera estar desincentivando el cumplimiento de los ASM de estas evaluaciones.

En el *gráfico 2.8*, se observa un declive en el porcentaje de ASM de desempeño atendidos durante 2012, cuando solo se cumplió con 28 por ciento (respecto al 78 por ciento de 2010). Se trata de una tendencia que se observó en todos los programas analizados (*véase gráfico 2.12*). Al respecto, el funcionario de Coneval “A” mencionó que la rotación de personal generada por el cambio de gobierno durante 2012 afectó considerablemente la capacidad de respuesta de las unidades responsables de los programas para dar seguimiento a las evaluaciones, en muchos casos por falta de experiencia o capacitación en el tema.

Por otro lado, el cumplimiento de los ASM de las ED también descendió entre 2007 y 2011, oscilando de 85 a 51 por ciento. Esta tendencia se observó en ocho programas, sin embargo, los que disminuyeron más dramáticamente fueron: Atención a Jornaleros Agrícolas (100 a 14 por ciento), Vivienda Rural (83 a 9 por ciento), y Abasto Social de Leche de LICONSA (100 a 37.5 por ciento).

(b) LA ATENCIÓN DE ASM POR TEMÁTICA

Los resultados obtenidos por temática muestran que los programas de vivienda o espacios públicos son los que reportan un porcentaje más alto de cumplimiento (69 por ciento), seguidos por los de desarrollo (67 por ciento), los productivos (61 por ciento), y por último, los de género (55.6 por ciento).

En el *Gráfico 2.12*, se puede apreciar que los porcentajes más bajos reportados durante el periodo tuvieron lugar en 2012, cuando los programas de desarrollo atienden solo 25.5 por ciento de los ASM comprometidos y los programas productivos 28.6 por ciento.

En un análisis más detallado, identifiqué que los ocho programas de desarrollo evaluados en 2012, mostraron un importante declive en el porcentaje de ASM atendidos. No obstante, los que obtuvieron los niveles más bajos fueron: Atención a Jornaleros Agrícolas (6.7 por ciento), 70 y Más (10.7 por ciento), y Desarrollo de Zonas Prioritarias (14.5 por ciento).

En el caso de los cuatro programas productivos evaluados en 2012, también observé una tendencia generalizada a la baja, donde los porcentajes de cumplimiento más bajos se registraron para los siguientes programas: Coinversión Social (23.3 por ciento), y el Fondo Nacional de Fomento a las Artesanías (23.8 por ciento).

(c) LA ATENCIÓN DE ASM POR PROGRAMA

En el periodo analizado, el promedio ASM atendidos por los programas de Sedesol fue de 65 por ciento.

Los programas que cumplieron con el mayor porcentaje de ASM entre 2007 y 2014 fueron: Rescate de Espacios Públicos (91 por ciento), y Seguro de Vida para Jefas de Familia (90 por ciento). Por otra parte, el programa que registró el menor porcentaje de ASM fue el de Subsidio a Jóvenes (17.5 por ciento) seguido del de Desarrollo de Zonas Prioritarias (47 por ciento).

REFLEXIONES FINALES

A partir de la información que se presenta en este capítulo, vale la pena resaltar algunos hallazgos tanto de la práctica como de la utilización de las evaluaciones a los programas de Sedesol.

En torno a la práctica evaluativa, hay tres aspectos que vale la pena enfatizar: i) la concentración de contrataciones; ii) la variación en los costos entre ED y EED; y iii) los cambios en la fuentes de financiamiento.

- i) Concentración en las contrataciones. Para las 111 evaluaciones analizadas se identificó la participación de 27 organismos evaluadores y 40 coordinadores. Destacando que EL COLMEX fue el organismo que concentró más evaluaciones (27 por ciento); y que solo tres coordinadores estuvieron a cargo de más del 35 por ciento de las evaluaciones: Israel Banegas (15 por ciento) y Martha R. Schteingart Garfunkel y Evelyne E. Rodríguez Ortega (10 por ciento cada una).
- ii) Variación en costes. En promedio, las ED fueron considerablemente más costosas y también registraron mayor variación entre el coste máximo y mínimo de las evaluaciones (hasta de 4 veces), en comparación con las EED cuyo coste varió solo hasta 1.6 veces.

En general, el costo promedio de las evaluaciones no excede a 0.05 por ciento del presupuesto de los programas. Sin embargo, de los 23 programas considerados, en las ED cinco exceden esa proporción: Fonart (0.9), PASPRAH (0.28 por ciento), PAJA y PAIMEF (0.25 por ciento), y PCS (0.22 por ciento). En el caso de las EED solo se registró una proporción más alta en PSPJ (0.21 por ciento) y Fonart (0.08 por ciento).

- iii) Fuente de financiamiento. Otro aspecto que vale la pena destacar es que Coneval dejó de asumir los costes de las EED. En la práctica, esta decisión no solo afecta las finanzas de las dependencias y entidades evaluadas sino que implica una menor participación de Coneval durante estos procesos y una menor distancia entre el evaluador y el evaluado, quien también sería el contratante.

En cuanto a la tendencia seguida en la utilización de las evaluaciones, en general, los datos mostraron que el aprovechamiento de los resultados ha sido considerable ya que más de 50 por ciento de las recomendaciones identificadas para el periodo fueron formalizadas como ASM por las unidades responsables de los programas, y que aproximadamente 65 por ciento de esos ASM fueron efectivamente implementados.

Cabe aclarar que las dependencias tienen la posibilidad de cancelar algunos de los ASM definidos, sin embargo, no encontré información sobre dichas cancelaciones. Por tanto, se debe considerar que el porcentaje de ASM que considero como no implementados puede incluir cancelaciones.

Por otra parte, la tendencia señalada no es constante a lo largo del periodo, se registra una disminución gradual tanto en la formalización como en la implementación de las recomendaciones. Los datos muestran que en 2014 los resultados fueron menos aprovechados respecto a las primeras experiencias en 2007, cuando recién se implementan los mecanismos que permitieron institucionalizaron esta práctica.

La tendencia observada podría ser natural en aquellos escenarios donde los programas que han mejorado gradualmente, cada vez requieren menos ajustes o éstos son de menor relevancia. Sin embargo, dados los últimos reportes que ha publicado Coneval⁴⁵ donde se reconoce que la pobreza se ha agudizado en el país (ascendiendo el porcentaje de pobres de 45.5 por ciento en 2012 a 46.2 en 2014), esto debería ser estudiado a más detalle en aquellos programas que específicamente están orientados a reducir este problema a fin de poder explicar este comportamiento.

⁴⁵ Para más información véase la Medición de la Pobreza de Coneval para el periodo 2010-2014 en: http://www.coneval.org.mx/Medicion/MP/Paginas/Pobreza_2014.aspx

CAPÍTULO 3

LA IMPORTANCIA DE LAS CARACTERÍSTICAS DE LAS EVALUACIONES Y DE LA INDEPENDENCIA DEL PROCESO EVALUATIVO EN LA UTILIZACIÓN

CAPÍTULO 3. LA IMPORTANCIA DE LAS CARACTERÍSTICAS DE LAS EVALUACIONES Y DE LA INDEPENDENCIA DEL PROCESO EVALUATIVO EN LA UTILIZACIÓN

En este capítulo analizo la relación existente entre factores (endógenos y exógenos) vinculados a la práctica evaluativa de Sedesol y la utilización de las recomendaciones emitidas por los evaluadores. El objetivo es identificar en qué medida estos factores influyen en que unas evaluaciones sean más utilizadas que otras. Para este fin, exploro 111 evaluaciones (de diseño y desempeño) realizadas a los programas federales de combate a la pobreza de la Sedesol, en el periodo 2007 - 2014.

Como ya mencioné al inicio de esta investigación (a partir de los postulados de la NGP), desde la década de 1990 las evaluaciones han recibido un gran impulso de numerosos gobiernos alrededor del mundo. Éstos atribuyen a la práctica evaluativa un papel estratégico, pues genera procesos de aprendizaje y de rendición de cuentas que favorecen tanto el mejoramiento como la democratización de la administración pública. No obstante, a lo largo de los años, la experiencia demuestra que los beneficios vinculados a esta práctica no siempre se analizan. Este hecho ha motivado la realización de numerosos estudios que pretenden explicar, en su mayoría, la dificultad o resistencia que los usuarios de la evaluación enfrentan para llevar a cabo las recomendaciones.

Para el análisis que efectué en este capítulo tomé en consideración dos trabajos de gran relevancia: Cousins y Leithwood (1986) y Balthasar (2009). En el primero, los autores identifican factores endógenos (aquellos vinculados a las características de la evaluación o a su proceso). Mientras que, en el segundo, Balthasar (2009) explora la relevancia de los factores exógenos (los arreglos institucionales).

En México, los estudios empíricos que abordan la relevancia de dichos factores para la utilización de las evaluaciones son, en su mayoría, cualitativos, y se apoyan en entrevistas. Por tal motivo, una de las aportaciones de este trabajo es realizar un análisis cuantitativo preciso a través del cual se explore la relación entre once factores (cinco endógenos y seis exógenos) y la utilización en las evaluaciones de Sedesol en estudio. Para cada grupo de factores llevo a cabo un análisis de regresión lineal multivariada, donde el porcentaje de recomendaciones atendidas por los programas es la variable dependiente.

El capítulo se compone de tres apartados. En el primero, señalo aspectos generales sobre la normatividad que rige la realización y la utilización de las evaluaciones de Sedesol. Este apartado permite conocer cuáles son las iniciativas adoptadas por el gobierno mexicano para incidir tanto en las características de los informes como en la implementación de las recomendaciones.

En el segundo, presento un análisis de regresión multivariada de mínimos cuadrados ordinarios (MCO) para determinar qué tan relevantes son cinco factores endógenos para la aplicación de las recomendaciones. Las variables empleadas son: la intensidad de evaluación en los programas, el número y la relevancia de las recomendaciones emitidas en los informes, la difusión de los resultados en medios electrónicos, y la procedencia del evaluador.

Por último, en el tercer apartado, también realizo un análisis de regresión multivariada de MCO, a través del cual exploro la relevancia de factores exógenos a partir de seis variables que reflejan la independencia del proceso evaluativo (o la distancia entre evaluador y evaluado): el actor que promueve la iniciativa, la unidad responsable de la supervisión, el origen del presupuesto, los términos de la contratación, la implementación de la evaluación y la divulgación de los informes.

3.1 ASPECTOS GENERALES SOBRE LA REALIZACIÓN Y LA UTILIZACIÓN DE LAS EVALUACIONES

En este apartado señalo, mediante la revisión de documentos oficiales, aspectos generales que rigen la realización y la utilización de las evaluaciones en Sedesol. Antes de ahondar en la relevancia que tienen determinados factores endógenos o exógenos (apartados dos y tres del presente capítulo) para la aplicación de las recomendaciones, consideré pertinente explorar las iniciativas que adoptó el gobierno mexicano para incidir tanto en las características de los informes como en el aprovechamiento de los resultados.

Cabe aclarar que los aspectos generales que presento a continuación no buscan explicar los factores que analizo más adelante, pero sí contextualizarlos. Para ello, busqué aclarar dos temas clave: qué criterios administrativos o legales se atendieron durante la realización de las evaluaciones que estudio, y cómo se promovió la implementación de las recomendaciones emitidas.

3.1.1. ¿QUÉ CRITERIOS SE ATIENDEN EN LA REALIZACIÓN DE LAS EVALUACIONES?

En México, tomando en cuenta la normatividad existente,⁴⁶ la realización de las evaluaciones está regulada por: primero, los Lineamientos Generales para la Evaluación de los Programas Federales de la Administración Pública Federal (Lineamientos Generales) que tienen como objetivo crear un sistema de evaluación y monitoreo que genere información oportuna y confiable para tomar decisiones; segundo, los Programas Anuales de Evaluación (PAE), mediante los cuales se establece un calendario de ejecución de las evaluaciones y se crea un vínculo con las actividades de programación y presupuesto del ejercicio fiscal correspondiente; y tercero, los Lineamientos de los Términos de Referencia para la Evaluación, los cuales son emitidos por Coneval con la finalidad de definir criterios homogéneos para la realización de los distintos tipos de evaluaciones que coordina, modificándolos eventualmente (Coneval, 2009).

A partir de la revisión del contenido de estos documentos, a continuación señalo qué criterios deben atender los dos tipos de evaluaciones consideradas en esta tesis: Evaluaciones de Diseño (ED) y Evaluaciones Específicas de Desempeño (EED).

(A) EVALUACIONES DE DISEÑO

En el caso mexicano, las ED se llevan a cabo bajo la supervisión de Coneval, siguiendo la metodología de la Matriz de Marco Lógico (MML). Estas evaluaciones se elaboran a partir de un cuestionario que busca resaltar la calidad en la definición de tres aspectos: los objetivos, la población objetivo, y la lógica horizontal y vertical del programa. En términos de planeación, ofrecen información de los objetivos nacionales y sectoriales a los cuáles contribuyen cada uno de los programas y la relación que guardan con otros programas federales que buscan resolver problemas afines (Coneval, 2015).⁴⁷

⁴⁶ La Ley de Desarrollo Social, la Ley de Presupuesto y Responsabilidad Hacendaria, el Decreto de Presupuesto de Egresos de la Federación 2007, el Decreto que establece las Medidas de Austeridad y Disciplina del Gasto de la Administración Pública Federal.

⁴⁷ Consultado en http://www.coneval.gob.mx/Evaluacion/MDE/Paginas/Evaluaciones_Diseno.aspx

Este tipo de evaluaciones se realiza tanto a programas nuevos como a aquéllos ya en operación. Para los que están en operación, con al menos dos años de existencia, el análisis del diseño se lleva a cabo cuando Coneval decide valorar la Consistencia y los Resultados.⁴⁸ Dicha evaluación es extensa y se conforma de seis apartados, sin embargo, para fines de este trabajo, sólo se emplea el primero que refiere a la lógica y la congruencia en el diseño de la intervención pública valorada (*véase cuadro 3.1*).

En el caso de los programas nuevos, los Lineamientos Generales señalan que es obligatorio que se les someta a una evaluación de diseño durante su primer año de operación, cumpliendo para ello con los mismos criterios establecidos en el primer apartado de *Diseño* de las evaluaciones de Consistencia y Resultados.

Cuadro 3. 1 Criterios para la realización de las Evaluaciones de Diseño, según los Términos de Referencia

Términos de Referencia Lineamientos Generales: Evaluaciones de Consistencia y Resultados	Antigüedad del Programa	Programas en Operación	Programas Nuevos
		<i>Numeral décimo octavo: a partir de dos años de existencia</i>	<i>Numeral vigésimo segundo: el primer año en operación</i>
	Apartados	i. Diseño	i. Diseño
		ii. Planeación estratégica	
		iii. Cobertura y focalización	
		iv. Operación	
		v. Población objetivo	
		vi. Resultados	

Fuente: Lineamientos Generales.

Durante la revisión de casos observé que, para el periodo 2007-2014, las evaluaciones de diseño realizadas a los programas de Sedesol se concentran principalmente en dos años: 2007 y 2011 (con 41.7 y 52.8 por ciento de las evaluaciones totales realizadas, respectivamente). Casi la mitad de las evaluaciones de diseño realizadas se llevaron a cabo en 2007 cuando Coneval inició sus actividades con el propósito de atender las inconsistencias en los programas sociales. Posteriormente, en 2011, se repite este

⁴⁸ Las evaluaciones de Consistencia y Resultados analizan sistemáticamente el diseño y desempeño global de los programas federales para mejorar su gestión y medir el logro de sus resultados con base en la matriz de indicadores (fracción I del numeral Décimo Octavo de los Lineamientos, 2007). Este tipo de evaluación, analiza la capacidad institucional, organizacional y de gestión de un programa, y puede efectuarse a partir de los dos años de la creación del programa. Aborda seis temas: Diseño, Planeación estratégica, Operación, Cobertura y focalización, Percepción de beneficiarios, Resultados (CONEVAL).

ejercicio para seguir mejorando la lógica interna de los programas, es decir, ayudarlos a que su esquema efectivamente contribuya a la solución del problema social para el cual fueron creados.

En los documentos consultados, no encontré información sobre los criterios que emplea Coneval para determinar cuándo se llevan a cabo las evaluaciones de Consistencia y Resultados. No obstante, en entrevista con el funcionario “A” de Coneval señaló que después de que se aplica una ED a los programas durante su primer año, con las evaluaciones de Consistencia y Resultado se valora nuevamente la lógica y la congruencia de los programas a finales de cada sexenio.

Términos de referencia para las ED

Con el propósito de regular su elaboración, los Lineamientos Generales especifican diez elementos que deben considerar este tipo de evaluaciones (capítulo II, numeral décimo octavo, fracción I):

1. Identificación del problema y si el diseño ayuda a solventarlo;
2. Contribución a objetivos estratégicos;
3. Evidencia de su contribución positiva a la consecución de su Fin y Propósito;
4. Explorar la lógica vertical y horizontal de la Matriz de Indicadores para Resultados (MIR);
5. Definición de la población potencial y objetivo, y los mecanismos de selección;
6. Justificación de la población potencial y objetivo;
7. Criterios y mecanismos aplicados para seleccionar las unidades de atención (regiones, municipios, hogares, o individuos, entre otros);
8. En su caso, la estructura del padrón de beneficiarios conforme a las disposiciones aplicables;
9. Reglas de operación o normatividad que explique su operación o funcionamiento;
10. Posible complementariedad o duplicidad con otros programas.

Los elementos previamente enlistados fueron tomados en cuenta por Coneval para emitir los términos de referencia. Cabe mencionar que para el periodo analizado

identifiqué dos versiones, los primeros términos de referencia estuvieron vigentes para las evaluaciones realizadas entre 2007 y 2009; mientras que la segunda versión emitida fue empleada en las evaluaciones de 2011 y 2013. Siguiendo los primeros términos de referencia emitidos, las evaluaciones de diseño realizadas entre 2007 y 2009 atendieron un cuestionario 34 preguntas (*véase cuadro 3.2*).

Cuadro 3. 2 Temas considerados y número de preguntas en los términos de referencia para la elaboración de las evaluaciones de diseño

Temas	Número de preguntas	
	2007-2009	2011-2013
Justificación del Programa	n.a.	3
Contribución a los Objetivos Estratégicos	7	3
Lógica Vertical y Horizontal de la MIR	15	11
Definición de Población la Potencial y Objetivo	6	3
Vinculación ROP	2	n.a.
Complementariedad o Duplicidad de Acciones	4	1
Total	34	21

* En los años 2010, 2012 y 2014 no se realizaron evaluaciones de diseño.

Nota: n.a. = no aplica

Fuente: Elaboración propia.

Por su parte, los términos de referencia para las evaluaciones de diseño de 2011 y 2013, reducen el número de preguntas a 21, además, introducen tres nuevos incisos sobre la justificación del programa y eliminan 17 en el resto de los apartados. En mi opinión, al disminuir el número de preguntas se afectó la profundidad del análisis, debido a que también se restringe la especificidad de las preguntas (para ver los cambios sufridos en la estructura de los informes *véase anexos*).

(B) EVALUACIONES ESPECÍFICAS DE DESEMPEÑO

De acuerdo con los Lineamientos Generales, las evaluaciones específicas son aquellas que se realizan mediante trabajo de gabinete y/o de campo para analizar aspectos específicos para la mejora de los programas. Dichas evaluaciones generan información útil, rigurosa y homogénea para los servidores públicos de las dependencias, unidades de evaluación y gobierno federal que toman decisiones en el nivel gerencial (Lineamientos Generales, 2007 y Coneval, 2005).

En este sentido, las EED son un tipo de evaluación específica⁴⁹ que valora de manera sintética la implementación de los programas sociales durante un ejercicio fiscal. Su finalidad es mostrar el avance en el cumplimiento de los objetivos y las metas establecidas por los programas mediante el análisis de indicadores de resultados, de servicios y de gestión; mediante el análisis de indicadores así como de la revisión de la síntesis de información entregada por las unidades responsables de los programas que se encuentra contenida en el Módulo de Información para la EED (Coneval, 2015).⁵⁰

El PAE 2008 precisa que los Programas presupuestales (Pp) que están sujetos a la EED, son aquellos supeditados a reglas de operación y que entregan subsidios. Durante el periodo analizado, los programas de Sedesol estuvieron sujetos a este tipo de evaluación en cinco años: 2008, 2009, 2010, 2012 y 2014.

Cabe destacar que originalmente este tipo de evaluaciones no fueron creadas para mejorar los programas directamente sino para informar las decisiones de SHCP y del Congreso. Pero debido a la visibilidad que adquieren los hallazgos de las evaluaciones a través de los medios de comunicación, los programas decidieron aprovechar la información para definir ASM (Thania de la Garza, 2015; y Entrevista a funcionario “A”).

Términos de referencia para EED

Para la elaboración de las EED, Coneval establece que se deben reportar datos a través de un informe y un reporte ejecutivo donde se aborden los siguientes temas:

- *Resultados.* Avance en la atención del problema o necesidad para el que fue creado el programa.
- *Productos.* Avance en la entrega de bienes y servicios a la población beneficiaria y el ejercicio el presupuesto.

⁴⁹ La Secretaría de Hacienda y Crédito Público reconoce ocho tipo de evaluaciones específicas: Obra Pública, Calidad de los Servicios, Percepción de los Beneficiarios, Específica de Desempeño, Inclusión de Perspectiva de Género, Perspectiva de Juventud y Discapacidad, Programas Desarrollo Forestal, y Fondo del Ramo 23. véase http://www.shcp.gob.mx/EGRESOS/sitio_pbr/evaluacion/modelos/eed_2008_shcp.pdf

⁵⁰ http://www.coneval.gob.mx/Evaluacion/MDE/Paginas/evaluacion_especifica_desempeno.aspx

- *Presupuesto.* Cambios en los recursos ejercidos por el programa en el ejercicio fiscal evaluado y anteriores.
- *Cobertura.* Definición y cuantificación de la Población Potencial, Objetivo y Atendida, así como la localización geográfica de la Población Atendida.
- *Seguimiento de Aspectos Susceptibles de Mejora.* Avance en los compromisos de mejora que definen los programas a partir de las evaluaciones.

Cuadro 3. 3 Apartados considerados en los términos de referencia para las eeds

Apartados	2008	2009	2010	2012*	2014
Datos Generales del Programa	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Resultados – Productos / Resultados y Hallazgos	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Evaluación de la Cobertura / Población y Cobertura	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Seguimiento a Aspectos de Mejora / Avance y Aspectos de Mejora	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Alineación al Plan Nacional de Desarrollo y Presupuesto			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Conclusiones	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Observaciones			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Valoración General de los Indicadores / Indicadores	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Información Adicional			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Datos de la Institución Evaluadora	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/> **		
Información de la Coordinación y Contratación				<input type="checkbox"/>	<input type="checkbox"/>
Glosario			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

* Los términos de referencia 2012 siguen vigentes por lo que fueron empleados para las EED 2014.

** En este año los “Datos de la Institución Evaluadora” son un sub-apartado de “Información Adicional”.

Nota: En los años 2007, 2011 y 2013 no se realizaron EEDs para los programas de Sedesol.

Fuente: Elaboración propia.

Estos criterios fueron retomados en las cuatro versiones de términos de referencia que emitió Coneval entre 2008 y 2014. En el cuadro 3.3, se pueden observar que los cambios más relevantes adoptados a lo largo de los años, que consisten en la incorporación de dos nuevos apartados referentes a la “Alineación al Plan Nacional de Desarrollo” (a partir de 2010) y a la “Información de la Coordinación y Contratación” (en 2012), donde además de mencionar los datos generales del evaluador se reporta el costo de la evaluación, la forma de contratación y la fuente de financiamiento. Este último punto es importante en términos de transparencia debido a que da mayor visibilidad al organismo evaluador y al proceso de contratación (para ver los cambios sufridos en la estructura de los informes véase anexos).

3.1.2. ¿CÓMO SE PROMUEVE LA IMPLEMENTACIÓN DE LAS RECOMENDACIONES?

Debido a que en el 2006 la implementación de recomendaciones fue un de los principales retos en el aprovechamiento de las evaluaciones, en 2008 Coneval, SHCP y SFP emitieron el Mecanismo de Seguimiento de los Aspectos Susceptibles de Mejora (de aquí en adelante, Mecanismo de Seguimiento) con el objetivo de promover su *utilización*.⁵¹

A través de este Mecanismo se verifica que, una vez concluidos los informes de evaluación, la unidad responsable de la operación del programa emita un oficio dirigido a Coneval, SHCP y SFP, estableciendo su posición institucional ante la evaluación. Una vez recibido, las tres instancias mencionadas seleccionan los ASM a acordar con los operadores para mejorar la gestión del programa. Posteriormente se establece un Convenio de Compromisos de Mejoramiento de la Gestión que el programa debe instrumentar (Myriam Cardozo, 2015), sin embargo, este no garantiza que la evaluación se utilice para mejorar los programas.

De acuerdo con Montiel, a pesar de que este instrumento normativo se diseñó, aparentemente, para resarcir la ausencia de reglas sobre el aprovechamiento de las evaluaciones, algunos funcionarios de Coneval y de otras dependencias perciben que su creación no responde a fines de fiscalización, ni para establecer su obligatoriedad, sino para inducir la utilización de las evaluaciones, aunque sin ofrecer incentivos. Por ello, Montiel (2011) afirma que podría considerarse un ejercicio de simulación, de cumplir por cumplir.

No obstante las limitaciones que presenta el Mecanismo de Seguimiento, las evaluaciones realizadas se han utilizado de varias formas: para afianzar programas frente a cambios en el gobierno, incidir en las negociaciones presupuestales, influir en el debate público sobre las decisiones de gasto, entre otros. En opinión de Cejudo (2011), la transformación de los hallazgos en acciones de mejora que contribuyan a mejorar los resultados de un programa, ha sido quizá la utilización de menor interés, pero la de

⁵¹ Entendida como la aplicación intencionada y directa de los hallazgos de las evaluaciones para promover mejoras en la toma de decisiones o en la implementación de los programas valorados.

mayor potencia. En este sentido, para Maldonado (2011, p. 229), el aprovechamiento observado, lejos de una adopción automática e irreflexiva de las recomendaciones, ha sido una apropiación constructiva a partir de conocimiento privilegiado sobre las restricciones internas y su aplicabilidad.

Cabe destacar que entre las dependencias y las entidades evaluadas por Coneval, Sedesol es la única que cuenta con un Sistema de Seguimiento a Evaluaciones Externas (Sisee), que se emplea para vigilar el cumplimiento de los compromisos asumidos y sus tiempos. Sin embargo, el Sisee sólo parece haber sido efectivo al inicio de su aplicación, ya que Cardozo (2009) identificó, a través de entrevistas a funcionarios, que gradualmente la magnitud de los compromisos ha disminuido que éstos se limitan a ser meramente formales (ej. trámites administrativos como el envío de oficios solicitados) ante el temor de que la Secretaría de la Función Pública (SFP) o la Auditoría Superior de la Federación (ASF) pudieran acusarlos de incumplimiento.

Cardozo (2009), también encontró que lo más frecuente es que los programas implementen las recomendaciones y den seguimiento al proceso, pero que su tratamiento es muy burocrático, lento y formal, y que se reporta en términos de “porcentaje de cumplimiento”, sin considerar su contribución al programa valorado. También, que la información que el Coneval publica en sus informes de seguimiento, en algunos casos, contrasta con los documentos de trabajo que publica cada dependencia o entidad (Montiel, 2011, p. 52).

En cuanto al aprovechamiento que los diputados hacen de las evaluaciones, Montiel encuentra que algunos de ellos ni siquiera saben que existen evaluaciones externas o que las cofunden con las auditorías realizadas por la ASF. Y, en casos extremos, no referidos a la experiencia de Sedesol, una de las entrevistadas afirmó que incluso hay informes (no de Coneval) que no han sido leídos (2011). Cardozo (2009), a su vez, identificó que entre el ejecutivo y el legislativo no había una adecuada coordinación que facilitara el desarrollo de la cultura evaluativa, y que cada uno de los poderes persigue su propio interés al tratar de contar con su sistema de evaluación.

Por otra parte, en entrevista, Alejandro González Arreola (Director General de GESOC) destacó la relación que ha tenido su organización con la Comisión de Presupuesto y Cuenta Pública del Congreso durante el periodo analizado, y el interés

que ha mostrado esta última por la información generada en las evaluaciones para justificar decisiones. Este es uno de los pocos casos en donde consta que el poder legislativo aprovecha los resultados.

En resumen, de acuerdo con Montiel (2011), el sistema nacional de evaluación muestra indicios de funcionalidad, pero no necesariamente de que cierre el ciclo de las políticas públicas con la aplicación de las recomendaciones emitidas por los evaluadores. Se reconoce que al inicio hubo un incentivo poderoso que logró la creación de entidades reguladoras capaces de coordinar los procesos; pero que ahora, en una segunda fase, la preocupación es saber si se utilizan y cómo se utilizan las recomendaciones de dichas evaluaciones.

3.2. LA RELEVANCIA DE FACTORES ENDÓGENOS: ¿LOS ATRIBUTOS DE LA EVALUACIÓN AYUDAN A EXPLICAR SU UTILIZACIÓN?

El objetivo de este apartado es explorar en qué medida atributos específicos de las evaluaciones realizadas a los programas de Sedesol ayudan a explicar el porcentaje de recomendaciones utilizado durante el periodo 2007-2014.

En trabajos previos, donde se analiza la relevancia de factores endógenos a partir de entrevistas a actores clave, los resultados apuntan a los siguientes atributos: la calidad de la evaluación (Cejudo, 2011; Montiel, 2011), las restricciones de tiempo (Montiel, 2011), la evaluabilidad del programa, la confiabilidad de la información empleada, la claridad y precisión de las recomendaciones (Cejudo, 2011), su factibilidad (Patiño, 2014), y la credibilidad del evaluador (Gris, 2014).

En esta tesis exploro la relevancia de estos cinco factores a través de seis variables que en su mayoría son numéricas y que no han sido exploradas, como: el *número de las recomendaciones*, las *citas en Google y Google Scholar* (la difusión de los resultados en medios electrónicos), el *número de evaluaciones acumuladas por programa*, la *procedencia del evaluador* y el *alcance de las recomendaciones*.⁵² Se trata de cinco atributos que se desprenden de factores endógenos señalados en las investigaciones realizadas por Cousins y Leithwood (1986), Mark y Henry (2004), y Johnson *et al.* (2009) (véase cuadro 3.4).

Cuadro 3. 4 Factores endógenos analizados

Factores	Variables	Tipo
i. Hallazgos	Recomendaciones emitidas	Numérica
ii. Amplitud de la difusión	Citas en Google	Numérica
	Citas en Google Scholar	Numérica
iii. Intensidad de la evaluación	Evaluaciones acumuladas entre 2002 y 2014	Numérica
iv. Credibilidad o renombre	Procedencia del evaluador	Categórica
v. Relevancia	Alcance de las recomendaciones	Categórica
La calidad de la evaluación se mide del 0 al 6, a más puntuación mayor calidad		

Fuente: Elaboración propia.

⁵² También contemplé incorporar al análisis la puntualidad con que se entregaron las evaluaciones pero no hay información disponible.

La primer variable, el *número de recomendaciones* emitidas en los informes, alude a los hallazgos de la evaluación y la elegí por su relevancia para la toma de decisiones. Asumo que a mayor número de recomendaciones, éstas son más específicas y de mayor utilidad para los tomadores de decisiones.

Las siguientes dos variables, el *número de citas en Google* y en *Google Scholar*,⁵³ permiten estimar el acceso que otros actores no vinculados al proceso evaluativo tuvieron a los resultados (la difusión). Para su cálculo, tomé en consideración los periodos oficiales que se señalan en los Programas Anuales de Evaluación para publicar los informes. En estos casos, asumo que, a mayor número de menciones en los buscadores de Internet, es mayor el conocimiento de la población sobre los hallazgos y, también, la politización del tema o la presión social para la atención de los problemas identificados en las evaluaciones.

Elegí la cuarta variable, el *número de evaluaciones acumuladas* por programa (o la intensidad de la evaluación), por su incidencia en la formación de una cultura evaluativa en las unidades responsables de los programas. Por tanto, asumo que, a mayor número de evaluaciones realizadas, es mayor el conocimiento acumulado en torno a esta práctica y menor la resistencia a la implementación de los resultados.

La quinta variable, la *procedencia del evaluador*, explora la posibilidad de que la implementación de las recomendaciones esté influenciada por el renombre del organismo evaluador o por su postura institucional sobre el tema y los hallazgos. Para la construcción de esta variable clasifiqué los diferentes tipos de organismos evaluadores en tres grandes grupos, según su naturaleza institucional: públicos, privados e internacionales.

Definí la sexta y última variable, el *relevancia de las recomendaciones*, tomando en consideración si el contenido de las recomendaciones emitidas en los informes brinda orientación suficiente, o no, para ayudar a los usuarios a solventar los retos o

⁵³ En Google se hace una búsqueda abierta donde se cuentan los resultados disponibles a partir del día en que se publica la evaluación hasta las fechas de entrega del siguiente ciclo presupuestario. En Google Scholar se contabilizan los documentos académicos donde se hizo mención a las evaluaciones para el año inmediato a la publicación del documento.

debilidades que se identificaron. Desde este enfoque, a mayor alcance de las recomendaciones mayor debe ser su utilización. Para esta variable establecí tres niveles de relevancia: alta, media y baja.

Además de los factores endógenos mencionados, incorporé al análisis estadístico dos variables que brindan otras posibles explicaciones respecto la implementación de las recomendaciones: el año en que se inician las evaluaciones y su costo.

Por una parte, incorporando al análisis el año en que se realiza la evaluación, busco controlar el efecto que pudieron tener las coyunturas económicas y políticas que experimentó el país durante el periodo. En este sentido, el evento más importante que espero captar mediante esta variable es la elección presidencial de 2012, cuando el país vivió la transición del Partido Acción Nacional (PAN) al Partido Revolucionario Institucional (PRI) después de 12 años. Esto es importante debido a que la formalización de la estrategia de evaluación emergió durante dos sexenios del PAN; por tanto, consideré pertinente observar si el regreso del PRI tuvo repercusiones en la utilización de las evaluaciones (*véase capítulo 2*).

Por último, incorporé al análisis el costo de las evaluaciones, con la finalidad de estimar en qué proporción los recursos financieros destinados a esta práctica determinan la implementación de los resultados. Asumo que, a mayor cantidad de recursos empleados, mayor es la atención que reciben los procesos evaluativos y la presión hacia los usuarios para utilizar las recomendaciones emitidas en el mejoramiento de los programas.

3.2.1. ANÁLISIS DE DATOS

En esta sección, primero describo los datos empleados en la investigación y después presento los resultados del análisis. Como mencioné previamente, exploré 111 evaluaciones realizadas de los programas de Sedesol entre 2007 y 2014; no obstante, en este caso, debido a la falta de información en algunas variables el número de observaciones disminuyó a 95.

Para iniciar, exploro si efectivamente existe una relación entre la variable dependiente y las independientes seleccionadas. En el cuadro 3.5 muestro el número de observaciones disponibles, la media, así como los valores mínimos y máximos.

Cuadro 3. 5 Valores medio, mínimo y máximo de las variables

	Obs	Media	Min	Max
Utilización de las evaluaciones (porcentaje)	99	65.11	0	100
Recomendaciones emitidas	111	7.23	1	21
Citas en Google	111	228.31	1	2860
Citas en Google Scholar	111	15.52	0	103
Evaluaciones acumuladas	106	6.39	0	12
Procedencia del evaluador	111	1.89	1	3
Relevancia de las recomendaciones	111	2.32	1	3
Año de inicio de la evaluación	111	2010	2007	2014
Costo de la evaluación (miles de pesos)	111	245.62	70.05	1,044.00

Fuente: Elaboración propia.

En el cuadro 3.6, se presenta un análisis de correlación entre las variables explicativas y la utilización de las evaluaciones y, en general, se aprecia una relación de moderada a débil. En este caso, las correlaciones más fuertes registradas respecto a la variable dependiente fueron: el número de evaluaciones acumuladas (-0.28) y las citas en Google (-0.19).

Los resultados negativos de las dos variables previamente mencionadas, parecen ser reflejo de una tendencia sostenida a la baja que se dio en la utilización de las evaluaciones. La cual advertí en el análisis introductorio del capítulo 2.

Por otra parte, se observa que en la mayoría de los casos existe una correlación de moderada a débil entre las variables independientes. Las correlaciones más fuertes fueron:

- El número de recomendaciones emitidas y su alcance (0.47). En la revisión de las evaluaciones pude observar que en la mayoría de los casos donde había más recomendaciones éstas son más detalladas y están orientadas a solucionar problemas específicos.
- El número de recomendaciones emitidas y las citas en Google Scholar (0.20). Es posible que para los académicos sean de mayor interés las evaluaciones con más recomendaciones, que como ya mencioné suelen brindar información más detallada sobre los problemas de los programas valorados, respecto a aquellas evaluaciones que emiten recomendaciones generales.
- El número de evaluaciones acumuladas y las citas en Google Scholar (0.19). Al igual que en el caso anterior, para fines académicos o de investigación, las evaluaciones ofrecen información valiosa que usualmente es de difícil acceso para los estudiosos de los programas. Por tanto, es entendible que a mayor número de evaluaciones también sea mayor el aprovechamiento que se hace de estas fuentes de información para la generación de conocimiento.

En términos generales, en esta primera aproximación, observé una relación muy débil entre los factores endógenos seleccionados y la utilización de los resultados. Sin embargo, la baja correlación observada entre los factores explicativos me permitió validar una buena selección de variables.

Es importante tener en cuenta que los valores de el cuadro 3.6 son solo correlaciones por pares y no enunciados causales. No se controla la relevancia de los factores que compiten; por tanto, se pueden enmascarar relaciones causales más complicadas, como el tamaño y la significancia estadística.

Cuadro 3. 6 Matriz de correlación de las variables principales

Número de Observaciones
99

	Utilización Evaluación	Recomendaciones Emitidas	Resultados Google	Resultados G.Scholar	Evaluaciones Acumuladas	Procedencia Evaluador	Relevancia Recomendaciones
Utilización de evaluaciones (porcentaje)	1						
Recomendaciones emitidas	0.0056	1					
Citas en Google	-0.1885	-0.0259	1				
Citas en Google Scholar	0.0653	0.1948	0.0717	1			
Evaluaciones acumuladas	-0.2795	0.0517	-0.0691	0.1908	1		
Procedencia del evaluador	-0.0694	0.0767	-0.1127	-0.0554	-0.069	1	
Relevancia de recomendaciones	0.0104	0.4699	0.1289	0.0139	-0.0408	-0.0238	1

Fuente: Elaboración propia.

Debido a las limitaciones de la matriz de correlación, también llevé a cabo un análisis bivariado empleando una regresiones de Mínimos Cuadrados Ordinarios (MCO), a fin de conocer la capacidad de las variables independientes para explicar la utilización. Los resultados se presentan en el cuadro 3.7.

En el análisis de regresión bivariado, el coeficiente de la variable de control *Año de inicio de la evaluación* fuera la más significativa para la implementación de las recomendaciones, con un R^2 ajustado de 0.59. Se trata de una variable categórica que muestra un coeficiente positivo muy alto en 2007 pero que a partir de 2009 se convierte en negativo y se mantiene así el resto del periodo; tendencia que se intensifica en 2012 y 2014 cuando registra un decrecimiento en la implementación de las recomendaciones que supera un 60 por ciento. La relevancia de la variable año no fue una sorpresa. En el capítulo 2 se observa una tendencia descendente en la utilización de las evaluaciones, a lo largo del periodo analizado.

Por otra parte, en el análisis bivariado, al igual que en la matriz de correlación, los resultados de los factores explicativos (evaluaciones acumuladas y citas en Google) son afectados por el descenso sostenido que se da en la utilización, a lo largo de los años. Como consecuencia, arrojan coeficientes negativos que establecen una relación contraria a la que anticipaba en esta investigación. Se trata de resultados que carecerían de sentido si no se toma en cuenta el contexto.

Otra variable control que obtuvo un nivel de significancia alto fue el *costo de la evaluación* (99 por ciento), sin embargo, tanto su coeficiente como su capacidad explicativa (R^2) fueron muy bajos (véase cuadro 3.7).

Una vez observada la relación entre las variables principales (con la matriz de correlación) y la capacidad de cada una para explicar la utilización de las evaluaciones (mediante la regresión bivariada), presento a continuación los resultados del análisis de regresión multivariado.

Cuadro 3. 7 Análisis de regresión bivariada para factores endógenos

	Coeficiente	Prob>F	R ²	R ² Ajustado	Núm. Casos
Recomendaciones emitidas	0.7871343	0.3736	0.0082	-0.0021	99
Citas en Google	-0.0128154*	0.0533	0.038	0.028	99
Citas en Google Scholar	0.0618048	0.7007	0.0015	-0.0088	99
Evaluaciones acumuladas	-2.89564 ***	0.0061	0.0781	0.0682	95
Procedencia del evaluador	-5.671941	0.4656	0.0055	-0.0048	99
i. Procedencia del evaluador		0.4776	0.0153	-0.0052	99
Variable omitida (Privados)	73.15267***				
Públicos	-9.863407				
Internacionales	.8906661				
Alcance de las recomendaciones	1.089646	0.8129	0.0006	-0.0097	99
i. Alcance de las recomendaciones		0.7600	0.0057	-0.015	99
Variable omitida (Baja)	67.55769 ***				
Media	-5.19055				
Alta	-.5536014				
Año de inicio de la evaluación	-9.234469***	0.000	0.4371	0.4313	99
i. Año de inicio de la evaluación		0.000	0.6224	0.5933	99
Variable omitida (2007)	93.89615 ***				
2008	1.578391				
2009	-30.84733***				
2010	-15.60615*				
2011	-17.64615**				
2012	-65.96538 ***				
2013	-34.31115**				
2014	-62.72815***				
Costo de la evaluación (pesos)	0.0000398***	0.0028	0.0887	0.0793	99

* p < 0.1; **p < 0.05; *** p < 0.01.

Fuente: Elaboración propia.

3.2.2. RESULTADOS: REGRESIÓN MULTIVARIADA DE MÍNIMOS CUADRADOS ORDINARIOS

Para verificar los supuestos a los que asocio los cinco factores endógenos⁵⁴ y las dos variables de control, previamente expuestas; presento cuatro modelos de regresión multivariada de Mínimos Cuadrados Ordinarios (MCO) (véase cuadro 3.8).

En el primer modelo, solo analizo los factores endógenos y, nuevamente, las variables *citas en Google* y *evaluaciones acumuladas* son las que mostraron una relación significativa con la utilización de las recomendaciones. Sin embargo, la capacidad explicativa de este modelo fue muy baja, con un R^2 ajustado de apenas 9 por ciento.

En el segundo, además de los factores endógenos, incorporé al análisis la variable de control *costo de la evaluación*. Esta arrojó una alta significancia estadística (90 por ciento de confiabilidad) y prácticamente duplicó la capacidad explicativa del primer modelo, alcanzando un 18 por ciento. Sin embargo, al igual que en los resultados de la regresión bivariada, el coeficiente de esta variable es muy pequeño. Lo que significa que cada peso adicional que se asigna a la evaluación tiene un efecto positivo apenas perceptible en la implementación de las recomendaciones. Esto podría ser relevante si tomamos en cuenta que hay una importante variación en los costos, los cuales oscilan entre 70,000 y 1 millón de pesos.

Para el tercer modelo, agregué el *año de inicio de la evaluación* como variable numérica.⁵⁵ Como era de esperarse, ésta obtuvo un coeficiente negativo con alta significancia estadística. Además, por primera vez, las *citas en Google Scholar* y los *evaluadores públicos* (una de las tres categorías de la variable *procedencia del evaluador*), mostraron significancia estadística. Cabe destacar que la inclusión de la variable año incrementó la capacidad explicativa del modelo a 50 por ciento.

⁵⁴ Los datos del factor endógeno “Procedencia de la evaluación” fueron los únicos que se analizan como variable categórica, mediante la cual se identifican tres tipos de evaluadores: Privado (1), Público (2) e Internacional (3).

⁵⁵ A pesar de que es más conveniente analizar la importancia del tiempo “año” como variable categórica, lo cual hago posteriormente en el cuarto modelo. En éste caso, lo analizo como variable numérica con el propósito de observar su relevancia en el modelo pero sin concentrar toda la significancia.

Por último, recodifiqué la variable *año de inicio de la evaluación*, de numérica a categórica. A partir de este ajuste, aumentó la capacidad explicativa del modelo a 60 por ciento. Los resultados obtenidos, al igual que en modelo anterior, muestran una relación negativa entre esta variable y la utilización de las recomendaciones. Sin embargo, permitió observar que dicha tendencia se acentuó a partir de 2012, cuando se alcanzaron coeficientes superiores a -70.

En este cuarto modelo, desaparece la significancia de la variable *costo de la evaluación*, que se observó en los dos modelos anteriores y, también, pierde relevancia la variable *evaluadores públicos* (categoría sobre la procedencia del evaluador).

¿Los atributos de la evaluación ayudan a explicar la implementación de las recomendaciones?

En el análisis de regresión multivariada encontré una relación muy débil entre la implementación de las recomendaciones y los cinco factores endógenos seleccionados. Solamente dos llegaron a mostrar valores significativos, en ambos casos negativos: las variables: *citas en Google y Google Scholar* y la *intensidad de la evaluación*.

No obstante, dicha relación se desvaneció cuando introduje la variable de control *año de inicio de la evaluación*. Esto me permitió confirmar que la relevancia de los factores difusión e intensidad de la evaluación, observado previamente en los cuadros 3.6 y 3.7, solo es reflejo de cambios propios de la práctica evaluativa que se intensifican a partir del año 2012, cuando se dio un cambio de gobierno en las elecciones presidenciales afectando, aparentemente, la utilización de los resultados.

Cuadro 3. 8 Análisis de regresión multivariada para factores endógenos

												Núm Obs
												95
Utilización	Modelo1			Modelo2			Modelo3			Modelo4		
	Coef.	Std Err.	P> t	Coef.	Std Err.	P> t	Coef.	Std Err.	P> t	Coef.	Std Err.	P> t
Recomendaciones emitidas	-0.3280	1.1292	0.7720	0.0862	1.0818	0.9370	-0.887168	0.8514123	0.3000	0.0856	0.8049	0.9160
Resultados en Google	-0.0152	0.0065	0.022**	-0.0138	0.0062	0.028**	-0.004572	0.0049728	0.3600	-0.006373	0.0045	0.1630
Documentos académicos en Google Scholar	0.2288	0.1603	0.1570	0.1494	0.1544	0.3360	0.22582	0.1205788	0.065*	0.2374	0.1165	0.045**
Evaluaciones acumuladas	-3.3276	1.0483	0.002***	-3.4599	0.9979	0.001***	0.050007	0.9047181	0.9560	0.3911	0.8297	0.6390
Procedencia del evaluador												
Evaluador Público	-11.8813	8.7209	0.1770	-12.2863	8.2954	0.1420	-15.46575	6.467457	0.019**	-10.2076	6.4885	0.1200
Evaluador internacional	-6.3365	19.6056	0.7470	-16.8712	18.9369	0.3750	-22.60534	14.75229	0.1290	-3.4114	13.8416	0.8060
Alcance de las recomendaciones	1.2875	5.0854	0.8010	2.2065	4.8452	0.6500	4.13462	3.77817	0.2770	2.7259	3.5399	0.4440
Costo de la evaluación				0.000041	0.0000	0.002***	0.00003	0.00001	0.007***	0.0000	0.0000	0.5600
Año de inicio de la evaluación							-9.615275	1.27263	0.000***			
2008										-5.071124	17.2130	0.7690
2009										-32.65561	17.3781	0.064*
2010										-22.2520	18.2061	0.2250
2011										-13.96789	8.7524	0.1150
2012										-73.94905	17.4271	0.000***
2013										-78.78453	25.1906	0.002***
2014										-68.27095	16.6892	0.000***
Constante	94.4828	15.3388	0.0000	81.4571	15.1492	0.0000	19394.5200	2556.2060	0.0000	98.3538	21.6785	0.0000

3.3. LA RELEVANCIA DE FACTORES EXÓGENOS: ¿LA INDEPENDENCIA DEL PROCESO EVALUATIVO DETERMINA LA UTILIZACIÓN DE LAS RECOMENDACIONES?

En este último apartado, exploro en qué medida la independencia del proceso evaluativo determina la utilización de las recomendaciones. Para ello, analizo seis factores exógenos o institucionales que emplea Balthasar (2009) en su investigación para determinar la cercanía entre evaluador y evaluado; en otras palabras, la capacidad que tienen los últimos para influir en los resultados de la evaluación, a partir del marco legal vigente.

Los seis factores exógenos que considero en este análisis son: el actor promotor de la evaluación, la unidad responsable de la supervisión, el origen del presupuesto destinado a la evaluación, los términos de contratación, la implementación de las evaluaciones, y la divulgación de los resultados (*véase cuadro 3.9*).

Cuadro 3. 9 Factores exógenos o institucionales analizados

Factores	Interrogante	Distancia	
i. El actor que promueve la iniciativa	¿la evaluación se detona por una unidad independiente de...?	Órgano desconcentrado o unidad operadora del programa	0
		La dependencia evaluada (Sedesol)	1
		Otra (Coneval)	2
ii. La unidad responsable de la supervisión	¿existe una unidad responsable del proceso evaluativo ajena a...?	Órgano desconcentrado o unidad operadora del programa	0
		La dependencia evaluada (Sedesol)	1
		Otra (Coneval)	2
iii.El origen del presupuesto destinado a la evaluación	¿cubre el costo otra fuente distinta a la instancia evaluada?	Órgano desconcentrado	0
		La dependencia evaluada (Sedesol)	1
		Otra (Coneval)	2
iv.Los términos de contratación de la evaluación	¿los contratos se asignan a través de una licitación?	No, Adjudicación directa	0
		No, Convenio nacional	1
		No, Programa de Cooperación Internacional	2
		Sí, Concurso	3
v. La implementación de las evaluaciones	¿la realiza un evaluador externo?	No, Interno	0
		Sí, Externo	1
vi.La divulgación de los resultados	¿existe un informe final público?	No	0
		Solo dependencia coordinadora (Coneval)	1
		Dependencia coordinadora y acceso parcial en la dependencia evaluada (Coneval y link en Sedesol)	2
		Dependencia coordinadora y dependencia evaluada (Coneval y Sedesol)	3

Fuente: Elaboración propia.

En estudios previos realizados por Balthasar (2009) y Cejudo y Montiel (2011) identifican la relevancia del marco normativo en los proceso de contratación y en la distinción de la capacidad del evaluador; y por su parte, Gris (2014) establece la relevancia de la divulgación.

Cabe aclarar que en este apartado no exploro factores exógenos o institucionales vinculados a ordenamientos legales que garanticen la atención de las recomendaciones. Si bien estos son importantes, no son el objetivo de esta investigación, debido a que ya existen numerosos trabajos donde se hace énfasis en su importancia así como en la necesidad de que estos sean creados o fortalecidos.

Los datos empleados en el análisis se obtuvieron mediante la revisión de 111 informes, lo cual me permitió identificar la distancia entre evaluador y evaluado en cada uno de los seis factores (*véase cuadro 3.9*). A continuación explico brevemente qué medí a través de dichos factores y cuáles son los supuestos que los vinculan con la implementación de las recomendaciones.

El primer factor (la distancia entre el actor que promueve la evaluación y la unidad responsable del programa valorado) considera si los evaluadores atienden las necesidades de información de los usuarios. En este caso, se asume que entre más lejos esté el actor que promueve la iniciativa de evaluación respecto a la unidad operadora del programa, mayor será esa distancia, así como la independencia con que los evaluadores podrán realizar su trabajo y emitir recomendaciones objetivas. Sin embargo, para algunos autores esto podría ser cuestionable debido a que una corriente que defiende las evaluaciones próximas a los usuarios (ej. evaluaciones internas), puede brindar información más útil a los operadores.

Para el caso mexicano, fue imposible estimar la relevancia de este factor debido a que no había variación en los datos. Por una parte, porque la evaluación de programas sujetos a reglas de operación es definida como obligatoria desde el Presupuesto de Egresos de la Federación del año 2000 y, por otra parte, debido a que Coneval es el único responsable de dar a conocer los programas que serán valorados cada ciclo presupuestal en el PAE (*Norma de difusión de los resultados de las evaluaciones*, 2013).

El segundo factor permite estimar la distancia entre la unidad que coordina las evaluaciones y aquella responsable de operar el programa. Para el análisis, asumo que la distancia es mayor si una unidad ajena al proceso es la encargada de iniciar la evaluación y supervisarla, que si los responsables del programa gestionan la evaluación por sí mismos.

En México, los Lineamientos para la evaluación establecen que las dependencias y entidades, sin modificar su estructura, deben designar una unidad administrativa ajena a la operación de los programas federales para coordinar la contratación, operación y supervisión de las evaluaciones. Dicha unidad debe ser la responsable de garantizar la calidad y cumplimiento normativo, así como de coordinarse con Coneval, en el ámbito de su competencia, para el buen desarrollo de todas las etapas del proceso de evaluación (Lineamientos, 2007).

En Sedesol, la denominada Unidad de Evaluación es la que cumple con las funciones señaladas por los Lineamientos, siendo la principal responsable de coordinar la implementación de las evaluaciones de esta secretaría. Sin embargo, en esta investigación, además, considero que, cuando Coneval financia las evaluaciones, este organismo desempeña un rol activo en la coordinación del proceso.

Retomando el trabajo de Balthasar (2009), establecí para el análisis de este factor tres niveles, donde: i) la distancia máxima se da cuando la unidad responsable de la evaluación es ajena a la dependencia o entidad valorada (en este caso Coneval o las coordinaciones nacionales); ii) respecto al nivel anterior, la distancia se reduce cuando la unidad coordinadora forma parte de la dependencia o entidad valorada (Sedesol) y, por último, iii) no existe distancia cuando es la misma unidad quien coordina y quien es responsable del programa (órgano desconcentrado o unidad operadora del programa).

El tercer factor, determina la independencia del proceso de evaluación a partir del origen del presupuesto. Para estudiarlo, analicé las diferentes fuentes de financiamiento y asumí que existe mayor distancia entre evaluadores y evaluados si los recursos no provienen de la entidad valorada.

En México, el Reglamento de la LFPRH, establece que el costo de la evaluación deberá cubrirse con cargo al presupuesto de la dependencia o entidad correspondiente. No obstante, los Lineamientos (2007) mencionan la posibilidad de que sea con cargo al Coneval, en el ámbito de su competencia. La decisión sobre quién cubrirá los costos de las evaluaciones se da a conocer, cada ejercicio fiscal, a través del PAE.

En esta investigación, las fuentes de financiamiento observadas fueron: Coneval, Sedesol, y los órganos desconcentrados de Sedesol. En este sentido, establecí tres niveles de independencia a través de los cuales sugiero que el presupuesto proveniente de Coneval es el que brinda más independencia al proceso, mientras que el presupuesto de órganos desconcentrados no brinda independencia.

En el último factor, la divulgación, analizo la independencia de las evaluaciones a través de los términos de contratación. Partiendo del supuesto de que la independencia es mayor si los contratos se designan a través de un proceso competitivo (invitación abierta o selectiva a licitar). Definí cuatro niveles de independencia para los términos de contratación: el primero, donde la selección por concurso brinda mayor independencia; el segundo, donde se toman en cuenta los programas de cooperación; el tercero, donde se consideran los convenios; y, el cuarto, donde ocurre la adjudicación directa, con la cual no habría independencia en el proceso.

La normatividad aplicable a cada ente público determina el procedimiento para los diferentes tipos de contratación para la evaluación: Adjudicación directa, Invitación a tres, Licitación Pública Nacional, Licitación Pública Internacional, y otro (Norma difusión, 2013). En el caso de Sedesol, la opción de Adjudicación directa de contratos es viable cuando el monto es menor a 250 mil pesos.

Ívico Ahumada, en entrevista, explicó que la asignación de contratos que en el capítulo 2 se observó concentrada en un reducido número de organismos evaluadores, se debe a que, a lo largo de los años, se ha logrado una mayor profesionalización de los evaluadores y a que los procesos de contratación a través de invitaciones a tres o por adjudicación directa están influidos en gran medida por el prestigio de los organismos evaluadores, especialmente en el caso de las contrataciones que realiza Coneval.

El quinto factor determina la independencia de las evaluaciones analizando su implementación. Observa si el proceso se lleva a cabo fuera de la administración (evaluador externo) o por personas involucradas a las iniciativas gubernamentales que son valoradas (evaluador interno).

Para el caso de México, tampoco fue posible analizar la relevancia de esta variable en la utilización de las recomendaciones debido a que los Lineamientos señalan que todas las evaluaciones se llevarán a cabo por evaluadores externo. En este sentido, la LFPRH precisa que se deberán realizar evaluaciones por conducto de instituciones académicas y de investigación u organismos especializados, de carácter nacional o internacional, que cuentan con reconocimiento y experiencia en las respectivas materias de los programas.

En el último factor, la divulgación, analizo la independencia del proceso evaluativo mediante el acceso público a las evaluaciones. Para ello, asumo que la distancia entre evaluadores y evaluados es mayor si se publica el informe final. Para los evaluadores, este hecho significa que deben hacer su trabajo disponible para la discusión de la comunidad académica (Wollmann 2003, 344). Además, es la única manera de permitir meta-evaluaciones que tienen como propósito último evaluar la calidad de las evaluaciones (ej. Widmer 1996).

Al respecto, los Lineamientos para la evaluación en México, establece que las dependencias y entidades deberán dar a conocer de forma permanente a través de sus respectivas páginas de Internet, en un lugar visible y de fácil acceso, los documentos y resultados de todas las evaluaciones externas de sus programas federales dentro de los 10 días hábiles siguientes de haber sido entregados a las instancias (numeral vigésimo sexto, Lineamientos, 2007). Sin embargo, en algunos casos no fue fácil encontrar los informes en los portales de Internet y encima, aunque en menor medida, los links de Sedesol no funcionaban.

Además de los seis factores exógenos o institucionales mencionados, incorporé al análisis estadístico dos variables de control que brindan otras posibles explicaciones sobre la utilización de las recomendaciones: el *año de inicio de la evaluación*, y el *tema del programa*.

Como mencioné en el apartado anterior, la variable *año de inicio de la evaluación* permite observar los cambios que ha sufrido la práctica evaluativa de Sedesol a lo largo del periodo, y especialmente los cambios políticos derivados de la elección presidencial de 2012.

Por otro lado, al incorporar como variable explicativa *el tema del programa*, asumo que existen diferentes niveles de complejidad en la implementación de las recomendaciones, según el problema social que se atiende. Para ello, clasifiqué los problemas sociales atendidos por los programas en cuatro categorías: desarrollo, proyectos productivos, vivienda o espacios públicos y género.

3.3.1. ANÁLISIS DE DATOS

En esta sección, presento una descripción inicial de los datos que empleé para el análisis. De un total de 111 evaluaciones revisadas, la información disponible me permitió examinar 99 observaciones.

Cuadro 3. 10 Análisis

	Observaciones	Casos únicos	Media	Mínimo	Máximo
Proporción de recomendaciones utilizadas	99	64	65.11	0	100
Actor promotor de las evaluaciones (omitida)	111	1	1.00	1	1
Unidad responsable de la supervisión (omitida)	111	3	0.95	0	2
Origen del presupuesto	111	3	1.28	0	2
Términos de la contratación	111	4	0.77	0	3
Implementación (omitida)	111	1	1.00	1	1
Divulgación	111	3	2.12	1	3
Año de inicio de la evaluación	111	8	2010	2007	2014
Tema del programa	111	4	3.01	1	4

Fuente: Elaboración propia.

Al igual que en el apartado anterior, antes de presentar los resultados de los modelos primero analizo la relación entre la variable dependiente y las independientes. En el cuadro 3.10 se puede observar el número de observaciones disponibles, la media, así como los valores mínimos y máximos.

En el cuadro 3.11, presento una matriz de correlación donde se observa una relación moderada entre la utilización de las evaluaciones y tres factores instituciones: la *Unidad responsable de la supervisión* (-0.37), el *Origen del presupuesto* (0.35), y la *Divulgación* (0.50).

Sin embargo, también se observa una correlación de moderada a débil entre variables independientes, entre la *Unidad responsable* y el resto de los factores exógenos analizados: *Origen del presupuesto* (-0.42), *Términos de contratación* (0.34) y, la *Divulgación* (-0.29).

Cuadro 3. 11 Matriz de Correlación

						Número Observaciones	
						99	
	Evaluaciones utilizadas %	Actor promotor	Unidad responsable	Origen presupuesto	Términos contratación	Implementación	Divulgación
Proporción de evaluaciones utilizadas	1						
Actor promotor	.	.					
Unidad responsable	-0.3735	.	1				
Origen del presupuesto	0.3505	.	-0.4214	1			
Términos de contratación	-0.0606	.	0.3383	-0.0245	1		
Implementación	
Divulgación	0.5026	.	-0.2881	-0.0875	-0.166	.	1

Fuente: Elaboración propia.

Adicionalmente, para conocer la capacidad explicativa de las variables independientes seleccionadas, llevé a cabo un análisis bivariado empleando una regresiones de MCO (véase cuadro 3.12). En este análisis, las variables fueron examinadas como numéricas y como categóricas. En su formato numérico, tres factores institucionales examinados (*Unidad responsable*, *Origen del presupuesto* y *Divulgación*) y una variable de control (*Año de inicio de la evaluación*) mostraron una alta significancia estadística, con una confiabilidad del 99 por ciento.

Al analizar los datos a través de variables categóricas, a excepción de la variable *Temática del programa*, la mayoría mostraron una significancia alta. En este sentido, el *Año de inicio de la evaluación* fue la más relevante con un R^2 ajustado 0.59; seguido de los *Términos de contratación* (0.33); la *Unidad responsable* (0.24), la *Divulgación* (0.24) y; por último, el *Origen del presupuesto* (0.12).

Cuadro 3. 12 Análisis de regresión bivariada para explicar la utilización de las evaluaciones a partir de los arreglos institucionales

	Coefficiente	Prob>F	R^2	Adj. R^2	Núm. Casos
Actor promotor	omitida	.	0	0	99
Unidad responsable de la supervisión ¹	-13.33144***	0.0001	0.1395	0.1307	99
i. Unidad responsable de la supervisión		0.000	0.2594	0.244	99
Cons (Órgano desconcentrado o unidad operadora del programa)	72.36949***				
Sedesol	12.1479*				
Coneval	-26.97651***				
Origen del presupuesto ¹	14.01094***	0.0004	0.1228	0.1138	99
Origen del presupuesto		0.0008	0.1387	0.1207	99
Cons (Órgano Desconcentrado)	52.24909***				
Sedesol	4.553556				
Coneval	26.00858***				
Términos de la contratación ²	-1.883073	0.5513	0.0037	-0.0066	99
i. Términos de la contratación		0.000	0.3548	0.3344	99
Cons (Adjudicación directa)	76.91635***				
Convenio	-38.52728***				
Programa de cooperación	16.64865				
Concurso	2.362115				
Implementación	omitida	.	0	0	99
Divulgación ³	23.42315***	0.0000	0.2526	0.2449	99
i. Divulgación		0.0000	0.2621	0.2467	99
Cons (Coneval)	34.51118***				
Coneval y link en Sedesol	30.5409***				
Coneval y Sedesol	48.64055***				
Año de inicio de la evaluación(2007-2014)	-9.234469***	0.000	0.4371	0.4313	99
i. Año de inicio de la evaluación		0.000	0.6224	0.5933	99
Cons (2007)	93.89615 ***				
2008	1.578391				
2009	-30.84733***				
2010	-15.60615*				
2011	-17.64615**				
2012	-65.96538 ***				
2013	-34.31115**				
2014	-62.72815***				
Temática del Programa ⁴	0.839407	0.7582	0.0010	-0.0093	99
i. Temática del Programa		0.6947	0.015	-0.0161	99
Cons (Vivienda o Espacios Públicos)	69.07929***				
Productivo	-7.949643				
Género	-13.46262				
Desarrollo	-1.757521				

¹ Unidad responsable u Origen del presupuesto: Órgano desconcentrado o unidad operadora del programa (0), Unidad de evaluación de Sedesol (1), Coneval (2).

² Términos de la contratación: Adjudicación directa (0), Convenio (1), Programa de cooperación (2) y, Concurso (3).

³ Divulgación: Coneval (1), Coneval y link en Sedesol (2), Coneval y Sedesol (3).

⁴ Temática del Programa: Vivienda o Espacios Públicos (1), Productividad (2), Género (3) y, Desarrollo (4).

* p < 0.1; **p < 0.05; *** p < 0.01.

Fuente: Elaboración propia.

En general, se observaron tendencias negativas no previstas en las hipótesis establecidas al inicio del análisis. En dos de los seis arreglos institucionales considerados (*unidad*

responsable y términos de contratación) observé que a mayor independencia en el proceso evaluativo menor implementación de los resultados; asimismo, las dos variables de control (*el año de inicio de la evaluación y el tema del programa*) obtuvieron coeficiente negativos.

En un análisis más detallado, identifiqué que la utilización de las evaluaciones fue menor cuando las evaluaciones eran supervisadas por Coneval (-26.98), cuando eran contratadas mediante convenio (-38.5), y durante los procesos implementados en 2012 y 2014 (-65.96 y 62.72, respectivamente).

Una vez observada la relación entre las variables principales (con la matriz de correlación) y la capacidad de cada una para explicar la utilización de las evaluaciones (mediante la regresión bivariada). A continuación, presento los resultados del análisis de regresión multivariado.

3.3.2. RESULTADOS: REGRESIÓN MULTIVARIADA DE MÍNIMOS CUADRADOS ORDINARIOS

En el análisis de regresión multivariada de MCO, se omitieron dos factores (el *actor promotor de la evaluación* y la *implementación*) ya que no hay variación en sus datos. Esto debido, por un lado, a que el Coneval es oficialmente el único actor que promueve la evaluación de los programas a través del PAE, y por otro, a que las evaluaciones son llevadas a cabo solo por evaluadores externos (véase cuadro 3.10).

A continuación, presento tres modelos a través de los cuales analizo en qué medida los factores exógeno o institucionales que terminan la independencia del proceso evaluativo son capaces de explicar la utilización de las recomendaciones.

En el Modelo 1, analicé la relevancia que tuvieron para la utilización de las evaluaciones los seis arreglos institucionales (como variables numéricas). Solo dos mostraron significancia estadística (el *origen del presupuesto* y la *divulgación*), pero con una capacidad explicativa (R^2 ajustada) que alcanzó el 39 por ciento.

En el Modelo 2, analicé nuevamente la relevancia de los arreglos institucionales pero ahora como variables categóricas. La capacidad explicativa del modelo se elevó a 51.6 por ciento y, cuatro variables mostraron significancia estadística respecto a su relación con la utilización de las evaluaciones. Por un lado, observé que los resultados fueron más utilizados cuando: la *unidad responsable de la supervisión* era Sedesol (con un coeficiente de 25.59), cuando el *origen del presupuesto* provenía de Coneval (coeficiente de 33.67), y cuando la *divulgación* era efectiva tanto en Coneval como en Sedesol (coeficiente de 26.78). Por otro lado, se arrojó una relación negativa cuando los *Términos de contratación* fueron a través de convenios (-21.29).

En el tercer y último modelo, integro al análisis las dos variables de control, lo que incrementó la capacidad explicativa (R^2 Ajustado) a 63 por ciento. En este caso, prácticamente todos los factores no omitidos presentan valores significativos. Aquellos que mostraron una relación positiva respecto a la utilización de las evaluaciones fueron: los procesos donde la *unidad responsable de la supervisión* fue Sedesol (coeficiente de 16.05) y Coneval (coeficiente de 23.22); cuando los *términos de contratación* fueron

por concurso (coeficiente de 19.41); cuando la *divulgación* de los informes se da a través de ambas instituciones (Coneval y Sedesol) y por último, en los procesos evaluativos de 2013 (coeficiente de 39.35). Cabe aclarar que las evaluaciones realizadas en 2013 correspondieron solo a dos programas de nueva creación.

Por otro lado, se observan dos factores donde se presentan casos de relación negativa con la utilización de las recomendaciones: cuando Coneval financia las evaluaciones (coeficiente de -32.77), en su mayoría EED; y en los procesos evaluativos de 2012 cuando se da el cambio de sexenio (coeficiente de -109.2).

¿La independencia del proceso evaluativo ayuda explicar la utilización de las evaluaciones?

En tres de los cuatro arreglos institucionales que empleé para el análisis, observé una relación positiva entre éstos y la utilización de las evaluaciones que fue considerablemente mayor cuando hubo distancia entre evaluadores y evaluados (*unidad responsable de la supervisión, términos de contratación y divulgación*). El único caso donde esto no se presentó fue en el *origen del presupuesto* donde se obtuvo un coeficiente negativo cuando Coneval asumió el financiamiento de las evaluaciones, lo que llama mucho la atención puesto que sería el actor más distante.

No obstante, el hecho es que era de esperarse una mayor utilización de las recomendaciones cuando Coneval financia los procesos evaluativos, ya que asume un rol más activo en la coordinación y es el actor más distante, es importante tomar en cuenta que en su mayoría este Consejo cubrió la realización de EED, las cuales no tienen como objetivo principal mejorar los programas sino informar las decisiones de actores como la SHCP y los miembros del Congreso. Esto podría ayudar a explicar el coeficiente obtenido.

Cuadro 3. 13 Análisis de regresión multivariada para factores institucionales

								Número de Observaciones 99
Utilización	Modelo1			Modelo2			Modelo3	
	Coefficiente	Error estándar	Probabilidad P> t	Coefficiente	Error estándar	Probabilidad P> t	Coefficiente	Error estándar
Actor promotor de la evaluación	(omitida)			(omitida)			(omitida)	
Unidad responsable de la supervisión	-3.461689	3.515538	0.3270					
Sedesol				25.5975	8.0663	0.002***	16.0595	8.9060
Coneval				13.6363	8.6998	0.1210	23.2224	10.6967
Origen del presupuesto	14.25126	3.604551	0.000***					
Sedesol				-3.7024	6.7660	0.5860	-1.9644	6.1712
Coneval				33.6756	8.8553	0.000***	-32.7778	19.4511
Términos de contratación	2.075529	2.628855	0.4320					
Convenio				-21.2953	8.6515	0.016**	-0.0280	12.6949
Programa Cooperación				8.0050	16.4236	0.6270	-0.4186	15.4564
Concurso				0.0560	7.5179	0.9940	19.4195	10.6207
Implementación	(omitida)			(omitida)			(omitida)	
Divulgación	24.09187	3.952954	0.000***					
Coneval y link Sedesol				5.0349	8.1382	0.5380	109.7380	30.9366
Coneval y Sedesol				26.7794	8.6890	0.003***	75.6179	17.8485
Temática del programa							3.7208	1.8210
Año de inicio de la evaluación								
2008							11.4563	20.5524
2009							-18.0109	20.3306
2010							-18.6626	17.5074
2011							-14.8065	9.6091
2012							-109.2081	31.1186
2013							39.3589	21.6984
2014							(omitida)	
Constante	-1.4492	12.5062	0.9080	36.8854	9.0858	0.0000	-6.8919	17.2661
Prob > F			0.0000	Prob > F			Prob > F	
R ²			0.4169	R ²			R ²	
R ² Ajustado			0.3921	R ² Ajustado			R ² Ajustado	

* p < 0.1; **p < 0.05; *** p < 0.01.

Nota: Las variables omitidas por el modelo son el año "2007" y los "Evaluadores Privados"

Fuente: Elaboración propia.

REFLEXIONES FINALES

Si bien en la literatura no existe un consenso sobre la capacidad explicativa que tienen las características de las evaluaciones y los arreglos institucionales en cuanto a la utilización de las evaluaciones, ni en el análisis de variables que sean comparables entre investigaciones, considero que es indispensable explorar la relevancia de algunos de estos factores con el propósito de lograr un mayor entendimiento sobre el reto que enfrenta la implementación de las recomendaciones debido a que sólo así será posible tener una idea de qué se debe hacer para influir hacia un mayor aprovechamiento de las evaluaciones.

En el análisis de las características de las evaluaciones, sin embargo, no encontré los resultados esperados y fue la variable de control *año de inicio de la evaluación* la que ayudó a explicar la relación negativa advertida entre algunos factores (*resultados en Google* y *el número de evaluaciones acumuladas*) y la utilización de las recomendaciones. En este sentido, el descenso registrado en la atención de ASM entre 2007 y 2014, resultó ser el aspecto más relevante para este análisis.

Una vez controlado el efecto del tiempo, pude apreciar que factores como la *intensidad de evaluación* y las *citas en Google*, no mostraban una relación significativa con la utilización de las evaluaciones.

Por otra parte, en tres de los cuatro arreglos institucionales que pudieron ser empleados para el análisis, si observé una relación positiva considerablemente mayor donde había distancia entre evaluadores y evaluados (*unidad responsable de la supervisión, términos de contratación y divulgación*). El único caso donde esto no se presentó fue en el *origen del presupuesto* que obtuvo un coeficiente negativo cuando Coneval asumió el financiamiento de las evaluaciones, lo que llama mucha la atención puesto que sería el actor más distante.

No obstante, si bien esperaba encontrar una mayor utilización de las recomendaciones cuando Coneval financió los procesos evaluativos, es importante tomar en cuenta que este Consejo pago principalmente la realización de EED, las cuales

no tienen como objetivo principal mejorar los programas sino informar las decisiones de la SHCP y el Congreso. Esto podría ayudar a explicar el coeficiente obtenido.

CAPÍTULO 4

LA UTILIZACIÓN INSTRUMENTAL DE LAS RECOMENDACIONES DERIVADAS DE LAS EVALUACIONES Y SU EFECTO EN EL MEJORAMIENTO DE LOS PROGRAMAS

CAPÍTULO 4. LA UTILIZACIÓN INSTRUMENTAL DE LAS RECOMENDACIONES DERIVADAS DE LAS EVALUACIONES Y SU EFECTO EN EL MEJORAMIENTO DE LOS PROGRAMAS

El objetivo de este capítulo es explorar las mejoras que ha generado la utilización de las evaluaciones en la Secretaría de Desarrollo Social (Sedesol). Para ello, analizo el efecto de las 111 evaluaciones (36 de diseño y 75 de desempeño) que se llevaron a cabo en esa Secretaría entre 2007 y 2014.⁵⁶

Actualmente, el meta-debate en torno al aprovechamiento de las evaluaciones se centra, casi exclusivamente, en cómo se utilizan sus resultados y si mejoran la manera en que el programa atiende el problema público. Es decir, hay un vacío en la literatura sobre el efecto de esta práctica en el mejoramiento de los programas y la solución del problema social que atienden. Por ello, el interés central de este capítulo es aportar evidencia sobre los efectos de la evaluación.

Existe un reducido número de investigaciones en torno a este tema, por la dificultad de obtener información sobre los cambios que genera la práctica evaluativa, así como porque es complejo asilar los efectos no vinculados a esta práctica que también influyen en los ajustes de los programas. Sin embargo, en años recientes esta corriente de investigación ha mostrado mayor desarrollo como describo a continuación.

Entre los trabajos más destacados se encuentran los de Kirkhanrt (2000) y Henry y Mark (2004), quienes proponen un marco de análisis a través del cual se enfatiza la importancia de estudiar el impacto de las recomendaciones atendidas, considerando incluso aquellos impactos no inmediatos o no intencionados que usualmente son ignorados por quienes analizan la manera en que se utilizan las evaluaciones (*véase capítulo 1*).

⁵⁶ Para este periodo, Coneval reporta un total de 134 evaluaciones de diseño y desempeño aplicadas a 27 programas presupuestarios de la Sedesol. En esta investigación solo se analizan 111 que cuentan con la valoración de GESOC: 36 de diseño y 75 de desempeño (*véase anexos*).

En el caso mexicano, los estudios que han abordado el efecto de la evaluación analizan sólo al impacto directo e intencionado de las recomendaciones. Cejudo y Maldonado (2011) observan, a través de diez “casos exitosos”, los productos obtenidos a partir del cumplimiento de los ASM. Patiño (2014) y Gris (2014), analizan los cambios en las Reglas de Operación de los Programas (ROP). En todos los casos, los hallazgos mostraron una mejora en los programas valorados, pero no se conoce su relevancia para el mejoramiento del desempeño o el cumplimiento de los objetivos.

Hoy en día, a pesar de que se están dando los primeros pasos para la generación de este tipo de evidencia, el desconocimiento sobre el efecto de las evaluaciones dificulta hablar sobre la conveniencia o la efectividad de esta práctica (más allá de la realización de ajustes en los programas). Asimismo, opaca el debate entre quienes critican⁵⁷ y quienes buscan justificar los recursos financieros y humanos que se emplean. En este sentido, la principal aportación de este capítulo es generar evidencia en torno al efecto que han tenido las evaluaciones aplicadas a Sedesol en el diseño y desempeño de sus programas.

El capítulo se compone de dos apartados. En el primero, examino la relación entre las recomendaciones de diseño emitidas y los cambios en las ROP de los programas, un efecto inmediato e intencionado de las evaluaciones. Para lograrlo, llevé a cabo una revisión documental de 36 Evaluaciones de Diseño (ED) y de 71 ROP, la cual me permitió comparar el antes y el después de las evaluaciones y también determinar si éstas habían tenido algún efecto.

En el segundo apartado, exploro la relación entre la utilización de 75 Evaluaciones Específicas de Desempeño (EED) y la estimación del Índice de Desempeño de Programas Públicos Federales (INDEP), un efecto no inmediato pero intencionado de esta práctica. Para ello, utilizo un modelo de regresión lineal

⁵⁷ Entre las principales críticas identificadas entre académicos y evaluadores se encuentran, por un lado, cuestionamientos sobre la utilidad de resultados generados bajo restricciones metodológicas, de tiempo, presupuestales o de información, definidas por la legislación vigente y por Coneval. Por otro lado, la necesidad de diferenciar claramente las tareas de monitoreo de las de evaluación, esto debido a que muchas veces las últimas proveen información que, en realidad, corresponden al sistema de monitoreo generarla en menor tiempo y a menor costo (para más información, véase Nielsen y Ejler, 2008).

multivariada, donde el INDEP de GESOC⁵⁸ es la variable dependiente, y el porcentaje de ASM atendido es la variable explicativa principal.

⁵⁸ La organización de la sociedad civil GESOC publica periódicamente dos índices que ha construido para valorar a los programas evaluados: el Índice de Desempeño de los Programas Federales (INDEP) y el Índice de Calidad del Diseño (ICADI).

4.1. EL EFECTO DIRECTO EN EL DISEÑO: CAMBIOS EN LAS REGLAS DE OPERACIÓN

El objetivo de este apartado es determinar qué proporción de cambios en la lógica interna de los programas de Sedesol es resultado de aplicar las recomendaciones de las evaluaciones. Para ello, examino la relación entre las recomendaciones emitidas en 36 ED y los cambios en las ROP de 21 programas de la Secretaría, entre 2007 y 2014.⁵⁹ Me interesa explorar si este resultado es un efecto directo e intencionado de esta práctica, para alcanzar los objetivos del programa.

El presente análisis es relevante si tomamos en cuenta que no todos los ajustes en los programas se originan por las evaluaciones y que actualmente no existe evidencia que permita diferenciar entre las mejoras que han sido inducidas por esta práctica y aquellas originadas por otros procesos que pudieron experimentar los programas de manera simultánea. Por lo anterior, consideré pertinente examinar en qué medida los cambios en las ROP se vinculan de manera directa con las evaluaciones.

A partir de las evaluaciones, determiné la proporción de cambios realizados a las ROP, e identifiqué dos etapas: en la primera, efectué una revisión documental de las ED que me permitió cuantificar el total de recomendaciones emitidas en los informes. Posteriormente, en la segunda etapa, confronté las recomendaciones con el contenido de 71 ROP publicadas antes y después de las evaluaciones.

En el Cuadro 4.1, muestro el número de recomendaciones identificadas (emitidas y atendidas) en las ED. Distingo aquellas que por su contenido son aplicables al mejoramiento de las ROP. Cabe aclarar que a diferencia de las EED, los informes de las ED no asignan un apartado para enlistar las recomendaciones de los evaluadores. Por tanto, para el análisis contemplo todas las que identifiqué en el texto.

⁵⁹ De acuerdo con la SFP, las ROP son un conjunto de disposiciones que precisan la forma de operar un programa, con el propósito de lograr los niveles esperados de eficacia, eficiencia, equidad y transparencia. Disponible en: <www.funcionpublica.gob.mx/scagp/dgorcs/reglas/index.htm>

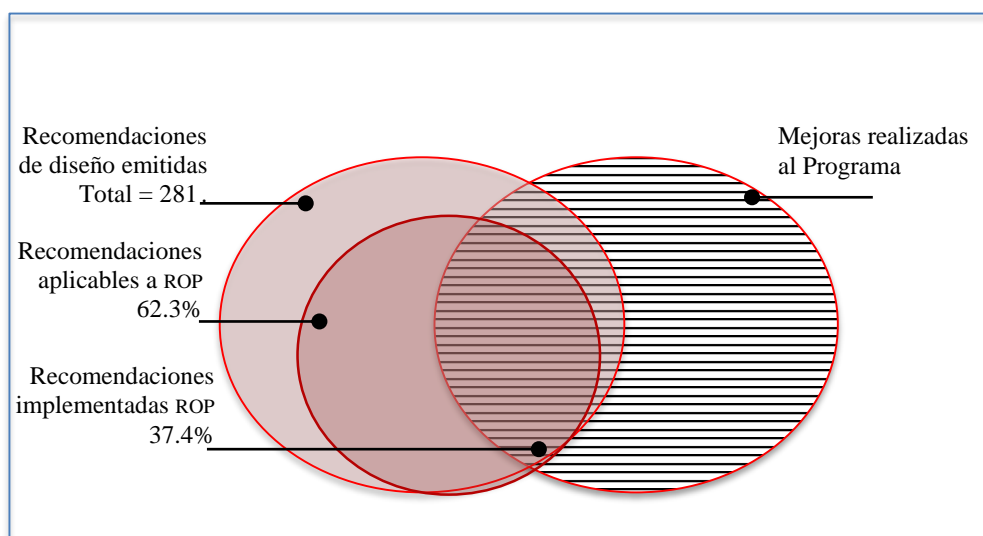
Cuadro 4. 1 Recomendaciones de diseño emitidas y atendidas

Recomendaciones Emitidas	Emitidas	Atendidas	Máximo	Media	Mínimo
Total identificadas	281 (100%)	-	19	7.81	1
Orientadas a las ROP	175 (62.3%)	105 (37.4%)	14	4.86	0

Fuente: Elaboración propia.

De un total de 281 recomendaciones que contabilicé en las 36 evaluaciones, 175 (62.3 por ciento) podrían incidir en la mejora de las ROP de los programas valorados y, de estas, fueron atendidas un poco más de la mitad (equivalentes a 37.4 por ciento del total de las recomendaciones emitidas) (véase *cuadro 4.1*). Destaca que este porcentaje de recomendaciones aplicado en ROP fue considerablemente menor que el porcentaje total de ASM que Sedesol reportó como atendidos (65 por ciento, véase *capítulo 2*).⁶⁰ Por lo anterior, asumo que una cantidad importante de recomendaciones de diseño ayudaron a mejorar los programas sin transitar por las ROP.

Ilustración 4. 1 Recomendaciones de diseño que fueron atendidas para mejorar las ROP, 2007 y 2014



Fuente: Elaboración propia con base en los informes de las ED publicadas por Coneval.

* Reglas de Operación (ROP).

Después de estimar en qué medida las recomendaciones de diseño se aprovecharon para mejorar las ROP, y con el propósito de identificar dónde tuvieron mayor impacto, a

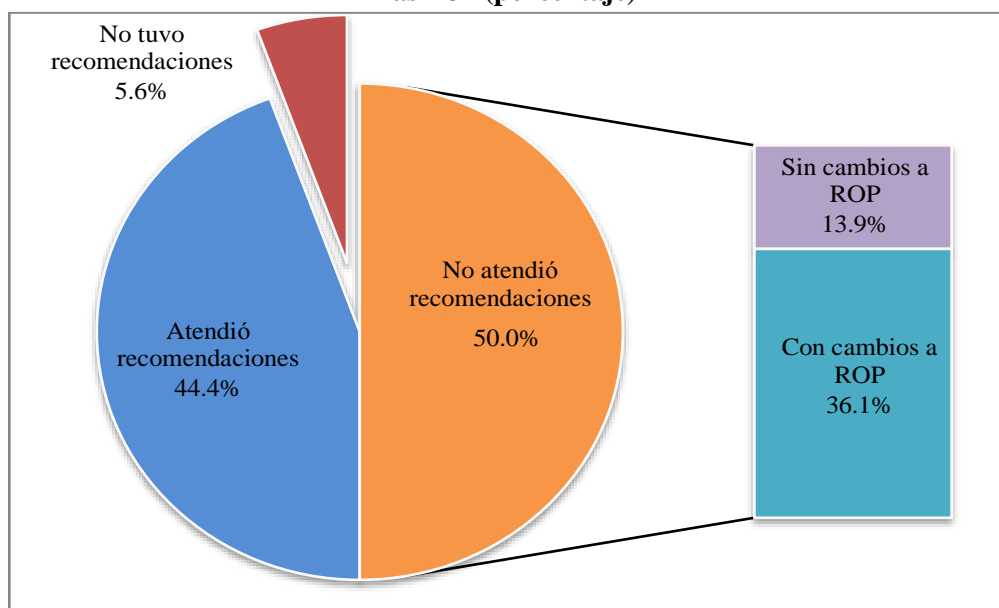
⁶⁰ Las cifras que presento en el capítulo 2 sobre la utilización de las recomendaciones de diseño, no son totalmente comparables con las que señalo en este capítulo, debido a que las últimas son producto de una revisión que hice para observar sólo las recomendaciones aplicables a las ROP de los programas, excluyendo cualquier otro compromiso administrativo derivado de las evaluaciones.

continuación analizo brevemente su implementación desde tres ángulos: (a) por cada evaluación realizada, (b) a través de los principales temas atendidos y (c) por programa evaluado. Posteriormente, con fines ilustrativos, empleo el Índice de Calidad de Diseño (ICADI) de GESOC para mostrar qué tanto varió el diseño de los programas valorados. Esta información me permite determinar si existe relación entre la utilización de las evaluaciones y las mejoras en el diseño.

(A) EVALUACIONES QUE DERIVARON EN CAMBIOS A ROP

En 34 de las 36 ED analizadas (94.4 por ciento) se emitieron recomendaciones orientadas a mejorar la lógica interna de los programas, la cual se plasma en las ROP. Sin embargo, solo 16 evaluaciones (44.4 por ciento) registraron cambios vinculados a dichas recomendaciones (*véase gráfico 4.1*).

Gráfico 4. 1 Porcentaje de evaluaciones que atendieron las recomendaciones orientadas a las ROP (porcentaje)



Fuente: Elaboración propia.

En aquellos casos donde no se atendieron las recomendaciones emitidas por los evaluadores (50 por ciento), observé dos escenarios: en el primero, las ROP tampoco sufren modificaciones (5 casos, equivalentes a 14 por ciento); mientras que en el segundo, los responsables de los programas sí realizan cambios a las ROP. Este último escenario fue el más importante debido a que concentró 13 casos (el otro 36 por ciento).

Fue una sorpresa descubrir que, en la mitad de los casos, las recomendaciones aplicables a las ROP no se implementaron. En general se espera que las ED influyan en la lógica interna de los programas. En especial porque muchos de ellos presentaban importantes retos en esta área (ej. población objetivo no definida, falta de diagnósticos y justificación para su operación, inconsistencia entre las actividades y su propósito, etc.). Al respecto, la Directora General Adjunta (DGA) de Evaluación de Coneval⁶¹ explica que las ROP son el documento que se emplea para auditar a todos los responsables de programas, siendo muy factible que cualquier cambio tenga implicaciones legales para ellos. Esto representa un gran desincentivo para la implementación de recomendaciones, sin importar qué tan pertinentes sean.

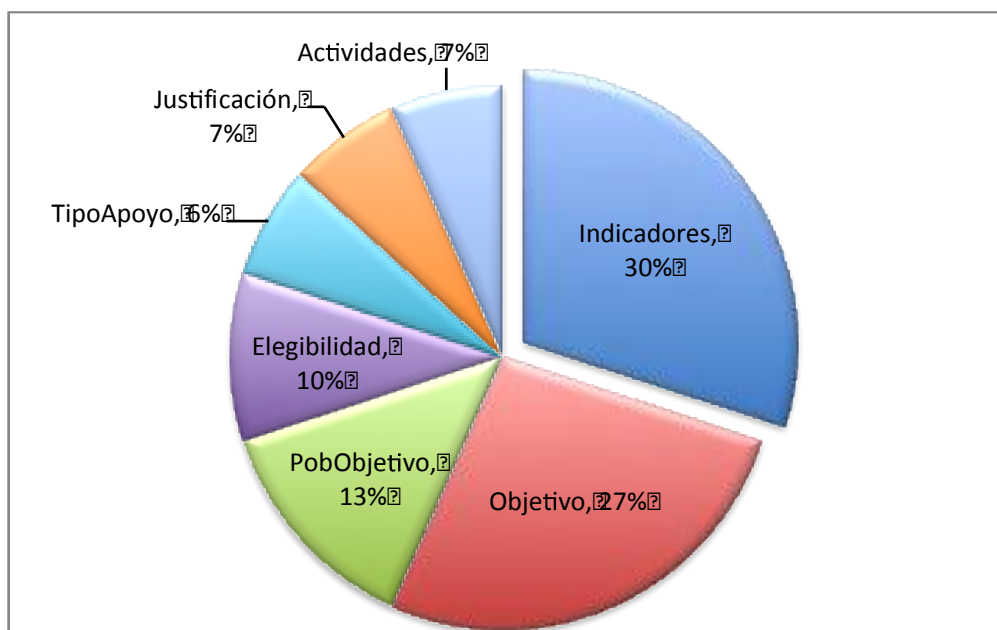
Por otro lado, la afirmación de la DGA de Coneval, torna más interesante el 36 por ciento de los casos donde los responsables de los programas optaron por ajustar las ROP por razones no asociadas con las recomendaciones. Convendría analizar en una investigación posterior la naturaleza de dichos ajustes, su relevancia, las implicaciones legales para los responsables de los programas y, sobre todo, qué procesos o actores influyeron para que se realizaran. En este trabajo, lo que pude observar es que, en su mayoría, los cambios se generan después de las evaluaciones de 2011.

(B) TEMAS DE LAS ROP MÁS IMPACTADOS POR LAS RECOMENDACIONES

Para el análisis comparativo de las ROP, antes y después de las evaluaciones, decidí observar los cambios realizados a través de once temas relativos a la lógica interna de los programas: el objetivo, el diagnóstico, la justificación, la población potencial, la población objetivo, la cobertura, los términos de colaboración intergubernamental, la elegibilidad de beneficiarios, el tipo de apoyo otorgado, los indicadores y las actividades (*véase gráfico 4.2*). Esto debido a que la estructura de las ROP es similar, pero presenta variaciones por programa y también a lo largo de los años.

⁶¹ De la Garza, Thania, Presentación: “¿Las evaluaciones se han usado para mejorar? El caso de las EED”. *Seminario Internacional de estudios empíricos sobre la utilización de la evaluación*. 24 de mayo 2016. Disponible en: <https://www.youtube.com/watch?v=nROCP3uIpVg&list=PLnojIu6biKTb4j7X9HG1BdG4uTWPePEVn&index=5>

Gráfico 4. 2 Temas que abordan las recomendaciones de diseño atendidas, 2007 y 2014 (porcentaje)



Fuente: Elaboración propia.

* El gráfico omite cuatro temas que no mostraron cambios: diagnóstico, población potencial, cobertura y colaboración intergubernamental.

En los hallazgos obtenidos identifiqué que los temas con mayor número de modificaciones fueron: los indicadores (9 casos), el objetivo del programa (8 casos) y, en menor medida, la población objetivo y la elegibilidad de los beneficiarios (4 y 3 casos respectivamente), véase gráfico 4.2. Se trata de cuatro temas importantes en la construcción de la Matriz de Indicadores para Resultados (MIR), cuya información permite el monitoreo de los programas. Resalta que la elaboración de la MIR y el seguimiento a sus indicadores está a cargo de la SHCP, la cual tiene capacidad de presionar a las dependencias evaluadas para realizar ajustes.

Por otra parte, no encontré ninguna modificación a partir de las evaluaciones fue en los siguientes cuatro temas: el diagnóstico, la población potencial, la cobertura y la colaboración intergubernamental. Al observar los ajustes totales realizados durante el periodo, vinculados o no a las evaluaciones, observé que estas cuatro variables también se encuentran entre las menos modificadas en general (véase gráfico 4.3).

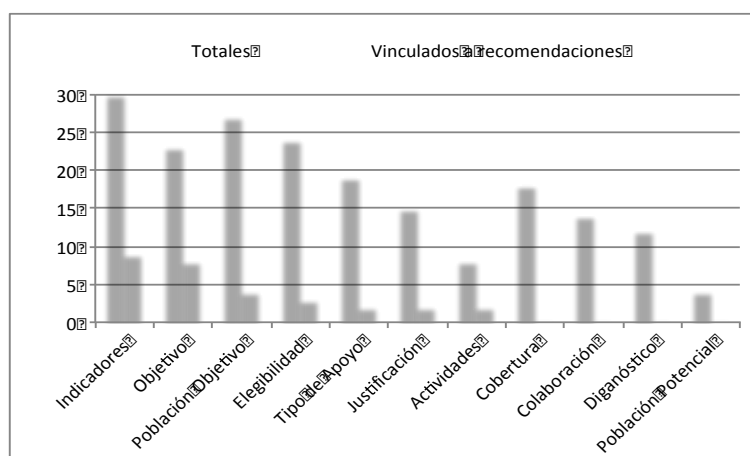
Al respecto, elaboro algunas apreciaciones que podrían ayudar a entender la falta y la resistencia a estos cambios:

Sobre el diagnóstico. Al iniciar las evaluaciones, la mayoría de los programas no contaban con un diagnóstico o no era adecuado. Además, los responsables de los programas no tenían incentivos para realizarlos, debido a que dichas iniciativas gubernamentales fueron creadas por inquietudes o decisiones de naturaleza políticas sin tener necesariamente un sustento científico. Por lo tanto, un diagnóstico adecuado muy probablemente pondría en evidencia la necesidad de un replanteamiento o recreación del programa. Además, para un programa en operación, este tema no es de gran relevancia puesto que no le impide desempeñar sus actividades ni condiciona el monitoreo que supervisa SHCP.

Sobre la población potencial. Al igual que con el tema anterior, cuando se iniciaron las actividades de monitoreo y evaluación, se reveló que los programas tenían poca o nula claridad respecto a su población objetivo, y los apoyos o recursos que ofertan. En consecuencia, enfrentaban serias dificultades para diferenciar entre sus poblaciones potencial y objetivo. Hoy en día, este problema se ha superado en la mayoría de los casos, pero la atención se ha centrado en definir la población objetivo, ya que es más importante para el ejercicio presupuestal y, como ya mencioné, para la MIR.

Sobre la cobertura y la colaboración intergubernamental. En el resto de los programas federales así como en los de Sedesol, estos dos temas responden a criterios políticos y están estandarizados. Por un lado, todos los programas tienen una cobertura nacional sin importar que su población objetivo se distribuya en un par de estados (ej. PAJA y 3X1) y, por otro lado, tanto su diseño como su operación está a cargo del gobierno federal, aunque estados y municipios sean responsables de tareas específicas. Cabe destacar que no identifiqué ni en las evaluaciones revisadas ni en los trabajos referentes al caso mexicano, un debate sobre estos dos temas o sobre las implicaciones de un modelo de combate a la pobreza centralizado.

Gráfico 4. 3 Número de casos donde se observaron cambios en las ROP por tema, 2007 – 2014*



Fuente: Elaboración propia.

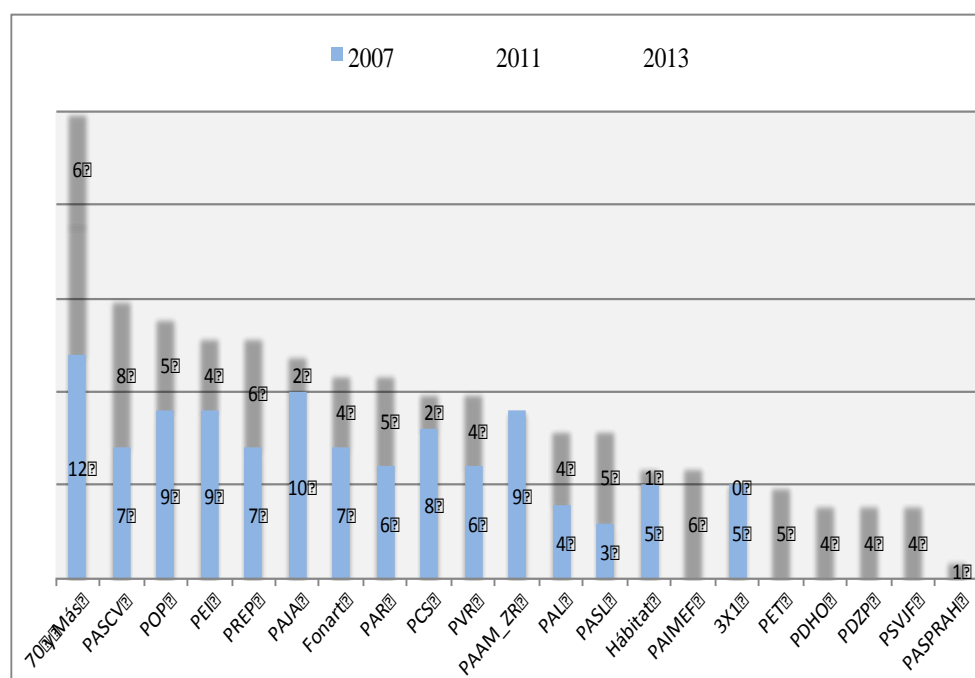
* Son los cambios observados en las ROP después de las 36 evaluaciones de diseño analizadas.

(C) PROGRAMAS CON MÁS AJUSTES EN SUS ROP A PARTIR DE LAS EVALUACIONES

Al explorar los cambios realizados por programa observé que, durante el periodo analizado, los 21 programas de Sedesol contemplados en este trabajo modificaron sus ROP por lo menos una vez (*véase gráfico 4.4*). Sin embargo, en ocho de estos programas (3X1, PAIMEF, PASPRAH, PDHO, PDZP, PET, PSVJF y PVR) no encontré ninguna relación entre los cambios y las recomendaciones (*véase gráfico 4.5*). Además, se reporta un bajo número de variables modificadas por programas a partir de las evaluaciones, dos en promedio.

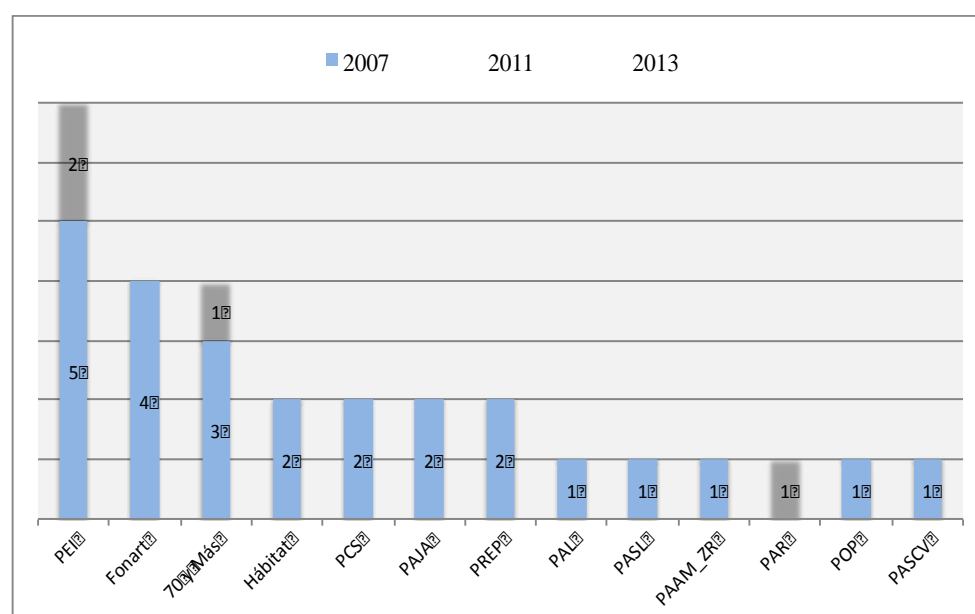
Otro aspecto que cabría analizar respecto a los cambios generados en los programas es que la mayor parte de ellos se realizaron en 2007, y disminuyeron considerablemente para 2011 a pesar de que en estos dos años se llevó a cabo un número similar de evaluaciones. Esta es una tendencia deseable y previsible en un proceso de mejora continua donde cada vez deberían necesitarse menos modificaciones, en particular en el diseño, aunque cabe la posibilidad de que el problema público haya cambiado.

Gráfico 4. 4 Cambios registrado en las reglas de operación (número de variables modificadas por año seleccionado)



Fuente: Elaboración propia.

Gráfico 4. 5 Cambios registrado en variables de las reglas de operación a partir de las evaluaciones (número de variables modificadas por año seleccionado)



Fuente: Elaboración propia.

En resumen, las recomendaciones de diseño tuvieron mayor impacto en 2007, sobre todo, en aquellas variables que son relevantes para la MIR. Además, identifiqué tres casos donde se atendieron más de dos recomendaciones por evaluación, todas en 2007: el PEI, realizada por el INSP (5 recomendaciones), el Programa de Fonart a cargo de Alternativas en Economía (4 recomendaciones), 70 y Más realizada por la Universidad de Chapingo (3 recomendaciones).

Por otro lado, observé que en ninguna de las dos evaluaciones realizadas a programas que abordan temas de género (PSVJF y PAIMEF) se atendieron las recomendaciones.

4.1.1. EL EFECTO DE LAS EVALUACIONES EN EL DISEÑO DE LOS PROGRAMAS

En este inciso, apoyándome en el ICADI de GESOC, exploro qué tanto han modificado su diseño los programas evaluados. Además, investigo si existe una relación entre los ajustes realizados a las ROP y la variación en el ICADI, sin que la intención sea señalar una causalidad. Al inicio de este trabajo se propuso estimar un modelo estadístico, sin embargo, debido al reducido número de observaciones disponibles, no fue posible y opté por un análisis exploratorio.

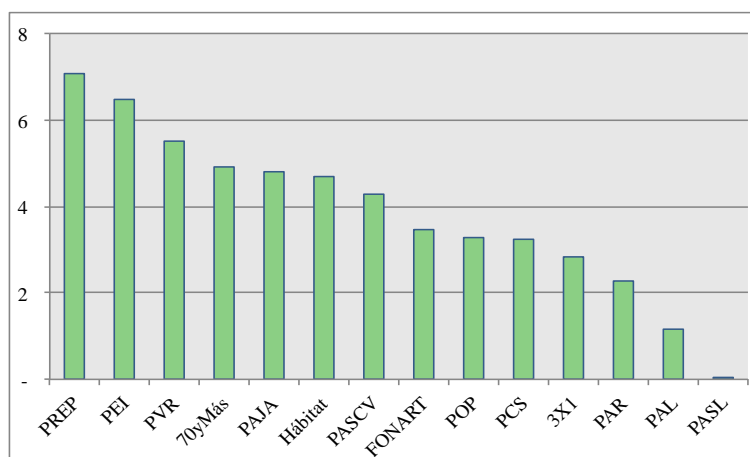
El ICADI se creó en 2007, basándose en la premisa de que un programa público que cuenta con “un diseño claro y coherente, con una visión estratégica de mediano plazo, una operación consistente y un perfil de orientación a resultados y a la ciudadanía”, tiene mayores probabilidades de tener un desempeño satisfactorio y de ofrecer mejores resultados vinculados a sus objetivos (ICADI, 2008).

Para la estimación de este índice, se emplean los resultados de las evaluaciones de diseño, a partir de los cuales se construyen tres subíndices que, de acuerdo con GESOC, resumen los factores clave de la calidad de diseño de un programa público: de diseño y alineación estratégica (objetivos, metas y alcances estratégicos del programa); de operación (previsiones institucionales y organizaciones vinculadas); y de orientación a resultados (previsiones para medir resultados e impactos o percepción y opinión de la población objetivo) (*véase anexo metodológico*).

Para estimar los cambios en el diseño de los programas comparé el ICADI en dos años (2007 y 2011). Sin embargo, la información disponible sólo me permitió realizar la estimación para 14 de los 21 programas. Los resultados obtenidos apuntan a una variación positiva en el diseño de todos los programas observados (*véase gráfico 4.6*). Destacando tres casos cuya variación fue mayor a cinco puntos, en una escala de 0 a 10: el Programa de Rescate de Espacios Públicos (PREP) que ascendió 7.05 puntos, el Programa de Estancias Infantiles para Apoyar a Madres Trabajadoras (PEI) con 6.49 y el Programa Vivienda Rural (PVR) con 5.5.

El PREP y el PVR se ubican entre los cinco programas con mayor costo promedio en sus evaluaciones (330 y 308 mil pesos, respectivamente). Mientras que el PEI reportó uno de los costos promedios más bajos (165 mil pesos).

**Gráfico 4. 6 Variación del valor del Índice de Calidad del Diseño (ICADI) 2007-2011
(Programas con más de una evaluación)**



Fuente: Elaboración propia.

Para explorar una posible relación entre la utilización de las evaluaciones y la mejora en el diseño de los programas valorados, decidí comparar los resultados de los gráficos 4.4 y 4.6, correspondientes a los cambios totales registrado en las ROP y la variación del ICADI. Los resultados del comparativo sugieren que aparentemente existe una relación entre estos temas.

De la misma manera, al comparar los gráficos 4.5 y 4.6, correspondientes a los cambios registrado en las ROP a partir de las evaluaciones y la variación del ICADI, también observé una posible relación entre los cambios en ROP derivados de las recomendaciones y la variación en el diseño de los programas, pero en una proporción mucho menor, destacando tres casos (3X1, PASL, y PVR) donde el mejoramiento en la calidad del diseño no implica una relación con las evaluaciones, debido a que no hay recomendaciones que se hayan atendido. De los tres casos señalados, llama más la atención el PVR pues, aunque no se atendieron las recomendaciones, presenta una estimación del ICADI que pasa de 3.5 a 9 entre 2007 y 2011.

¿La implementación de los ASM ayuda a explicar los cambios en el diseño de los programas?

La evidencia sugiere que puede existir una relación entre las recomendaciones implementadas y las mejoras en el diseño de los programas. Sin embargo, el efecto observado es limitado y llama la atención una proporción considerable de cambios en ROP no vinculados a las evaluaciones que valdría la pena estudiar en otra investigación.

Además, el efecto de las evaluaciones en el diseño de los programas parece ir en descenso, en la medida en que la lógica interna alcanza las características adecuadas. En este sentido, el mayor efecto se observó en las evaluaciones de 2007 y, en especial, en las variables más importantes para la construcción de la MIR.

4.2. EL EFECTO NO INMEDIATO EN EL DESEMPEÑO ¿LAS RECOMENDACIONES AFECTAN EL INDEP DE LOS PROGRAMAS?

Actualmente, existen altas expectativas sobre la capacidad que tiene la evaluación para mejorar la gestión pública a través de procesos de retroalimentación y mejora continua, aunque falta evidencia empírica que las sustente. De este modo, la principal aportación de este apartado es explorar el efecto de las evaluaciones a través de un análisis cuantitativo que me permita estimar en qué medida la utilización de las EED ayuda a mejorar el desempeño de los programas de Sedesol para el periodo 2008-2014 (un efecto intencionado, pero no inmediato, de las evaluaciones).

Cabe señalar que para el análisis cuantitativo, utilicé información disponible (oficial y generada por GESOC) que hasta el momento no había sido aprovechada para estos fines. De fuentes oficiales, empleé el porcentaje de ASM atendido por los programas que Sedesol emite anualmente en los informes de seguimiento (*véase capítulo 2*). De GESOC, rescaté el Índice de Desempeño de los Programas Federales (INDEP) que esta organización civil publica periódicamente para conocer la capacidad de los programas evaluados para alcanzar sus objetivos.

Este apartado se divide en dos secciones. En la primera, a manera de introducción, señalo aspectos generales sobre el INDEP, la variable dependiente del modelo de regresión que presento en el inciso 2.2 de este capítulo. Explico brevemente qué es, cuáles fueron los principales resultados obtenidos en su estimación durante el periodo y cómo se comporta este índice en los programas valorados. En este apartado yo no presento un análisis introductorio sobre el porcentaje de ASM atendido, principal variable explicativa del modelo, porque en el capítulo 2 se exploró de manera amplia.

En el segundo apartado, para analizar la relación entre la utilización de las evaluaciones y el desempeño de los programas utilicé un modelo de regresión lineal multivariada de MCO donde, como mencioné, el porcentaje de ASM atendido por los programas ($t-1$) es la variable explicativa principal, y el INDEP de GESOC es la dependiente.

4.2.1. EL ÍNDICE DE DESEMPEÑO DE LOS PROGRAMAS PÚBLICOS FEDERALES (INDEP)

GESOC⁶² creó este índice en el segundo ciclo de evaluación 2008-2009, mediante la sistematización y el análisis de los resultados de las EED efectuadas a programas federales sujetos a reglas de operación y otros de subsidio. El INDEP se estima periódicamente para analizar y desagregar los elementos que inciden en que un programa obtenga resultados, así como las interrelaciones existentes entre esos elementos.

Es importante aclarar que este índice no presenta una valoración del equipo investigador de GESOC, sino que recoge los resultados que las dependencias o entidades ejecutoras reportan a los evaluadores externos a través de indicadores y metas de desempeño por los ejecutores (GESOC, 2007 y 2008). Representa un esfuerzo por transparentar la denominada “caja negra” gubernamental que, en el mejor de los casos, permite conocer los insumos que utiliza el gobierno para alcanzar ciertos fines, pero no la forma en la que se procesan y transforman para producir los resultados o impactos deseados (INDEP, 2009).

Cabe señalar que no siempre es posible para GESOC estimar este índice por problemas de transparencia pues en numerosos casos, los responsables de los programas no entregan la información que los evaluadores solicitan.

La metodología elegida para elaborar el índice deriva del primer análisis a las EED en 2008, donde GESOC identificó tres problemas sistemáticos: la selección inadecuada de metas (*targeting*); la falta de información en los indicadores y la falta de un modelo de desempeño (debilidad en el sistema de indicadores).

Por lo anterior, la construcción del INDEP parte de una premisa simple: “todo programa gubernamental que se precie de contar con un buen desempeño debe contar

⁶² Organización de la sociedad civil creada en 2006 con el propósito de investigar, monitorear y evaluar el desempeño de instituciones públicas, organizaciones civiles y empresas de alto impacto social para incidir en la producción de resultados de valor público.

con al menos los siguientes atributos” (GESOC, 2008; *para más información véase anexo metodológico*):

- Un marco de evaluación comprehensivo y completo que permita analizar los avances de un programa en toda la cadena lógica del desempeño (insumos, procesos, productos, resultados e impactos).
- Metas claras y exigentes que permitan medir el esfuerzo potencial de una institución pública.
- Indicadores que deben mostrar avances significativos en la consecución de los objetivos de los programas.

Una vez creado el INDEP, GESOC trabajó en mejorar su estimación implementando algunos ajustes, entre los que destacan: incorporar al método de cálculo la calidad del diseño y cumplir con la cobertura de la población potencial. Sin embargo, las mejoras mencionadas no han sido suficientes para evitar que GESOC reciba severas críticas e académicos y funcionarios vinculados a la práctica evaluativa, los cuales ponen en duda la metodología empleada para estimar el INDEP y la confiabilidad de sus resultados que muchas veces exhiben el mal desempeño de programas y dependencias gubernamentales.

A pesar de las críticas dirigidas a GESOC, esta organización civil ha ganado prestigio entre la opinión pública e incluso ha sido galardonada por organismos nacionales e internacionales por su labor (Banco Mundial en 2012⁶³ y Coneval en 2015⁶⁴). Además, conviene recordar que este es el único esfuerzo realizado para estimar la capacidad de los programas de lograr sus objetivos, permitiendo medir y comparar las mejoras que experimentan en el tiempo.

⁶³ En 2012, se le otorga el *Premio a la Innovación en Transparencia para la Mejora de la Gestión Institucional*, convocado por el Banco Mundial, la ASF, el Instituto Federal de Acceso a la Información y Protección de Datos, el INAP y la SFP.

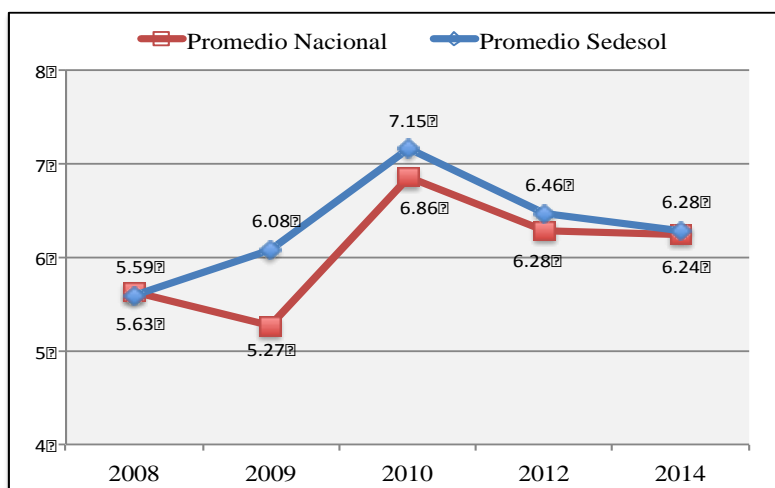
⁶⁴ En 2015, recibe un reconocimiento de Coneval por *Buenas Prácticas de Actores Sociales en el uso y generación de información para la mejora de políticas públicas y rendición de cuentas*. Específicamente por el desarrollo del INDEP, consolidado como la mejor herramienta para la medición del desempeño de los programas y acciones federales de desarrollo social.

Después de examinar los índices de GESOC y su método de cálculo, consideré que proveen información relevante y confiable que vale la pena aprovechar para los fines de esta investigación.

(A) EVOLUCIÓN DEL INDEP

En términos generales, al analizar el INDEP de los programas de Sedesol entre 2008 y 2014, observé que su desempeño mejoró a partir de las primeras evaluaciones, pero que esta tendencia pierde fuerza a partir de 2012. En este caso, a pesar de que en el capítulo 2 también identifiqué un retroceso en la utilización de las evaluaciones a partir de ese año, me resulta imposible determinar en esta tesis hasta qué punto la tendencia observada se debe a cambios en la Administración Pública Federal (APF), derivados de las elecciones presidenciales que dieron lugar a un cambio en el partido político en el poder, o a ajustes que realizó GESOC al método de cálculo del INDEP.⁶⁵

Gráfico 4. 7 Comparación de los índices de desempeño de los programas de Sedesol respecto al resto de los programas federales evaluados, 2008-2014



Fuente: Elaboración propia con datos de GESOC.

En el Gráfico 4.7 se observa que Sedesol tuvo un crecimiento sostenido entre 2008 y 2010, posicionando a sus programas por encima del promedio de todos los federales evaluados. No obstante, a partir de 2012, esta brecha se reduce y en 2014 la valoración

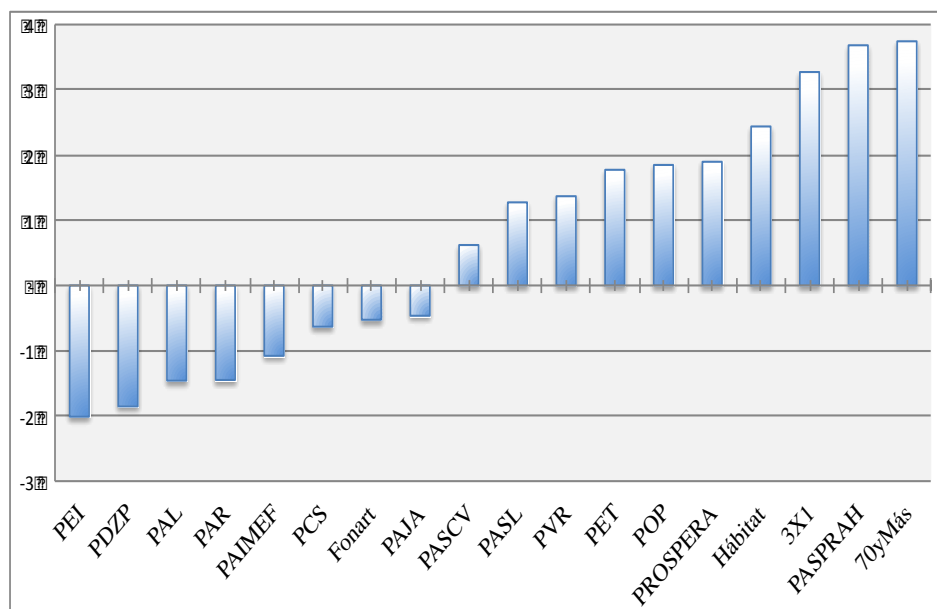
⁶⁵ En 2011, GESOC incorporó al método de cálculo del INDEP un nuevo subcomponente que permite considerar la calidad de diseño de los programas en la estimación de este índice (véase anexos).

promedio nacional y de Sedesol fue prácticamente la misma. Más aún, esta Secretaría tuvo un retroceso en la efectividad de sus programas.

Por otro lado, al analizar la variación que experimentaron los programas evaluados en su desempeño (véase gráfico 4.8), observé que se generaron experiencias positivas y negativas. Por un lado, entre las experiencias positivas más importantes se encuentran los programas con mayor número de evaluaciones y relevancia política, ej. 70 y Más, Hábitat y PROSPERA.

70 y Más atiende a adultos mayores en situación de vulnerabilidad y es un programa exitoso que pasó, en pocos años, de ser una iniciativa estatal a un programa nacional. Hábitat, por su parte, fue de los primeros programas en aplicar la focalización territorial de recursos, a través de los denominados polígonos hábitat constituidos por zonas urbanas marginadas. PROSPERA, por último, es un programa icónico en el combate a la pobreza que ha operado con un enfoque integral desde la década de los noventa en México bajo diferentes nombres.

Gráfico 4. 8 Variación del INDEP en cada programa de Sedesol evaluado, 2008-2014



Fuente: Elaboración propia con datos de GESOC.

En las experiencias negativas, donde los programas no incrementan sino que retroceden en su INDEP durante el periodo 2008-2014, identifiqué programas como PEI, PDZP y PAL. Llama la atención el caso del PEI porque en el apartado anterior observé que realizó

numerosos ajustes en su diseño a partir de las recomendaciones, por lo cual se esperaba que estos cambios hubiesen ayudado a su desempeño.

Entre los programas donde observé un decrecimiento del INDEP, hay tres operados directamente por Sedesol (PEI, PDZP y PAJA) y cinco por órganos desconcentrados (PAL, PAR, PAIMEF, PCS y Fonart).⁶⁶

⁶⁶ De un total de ocho programas que son operados por órganos desconcentrados cinco registran un retroceso en el INDEP y los tres restantes (PASCV, PASL y PVR) reportan los niveles más bajos de crecimiento.

4.2.2. CAMBIOS EN EL INDEP A PARTIR DE LA UTILIZACIÓN DE RECOMENDACIONES

En este apartado analizo la relación entre las recomendaciones atendidas de 75 EED y los cambios en el desempeño de 21 programas de Sedesol. El objetivo es explorar un efecto no inmediato pero intencionado de las evaluaciones (mejoras en el desempeño de los programas) que ha ayudado a justificar esta práctica desde su inicio. Sin embargo, en México, se presta poca atención cuando se habla de los beneficios obtenidos a través de las evaluaciones. En este apartado utilizo un modelo de regresión lineal multivariada donde la variación del INDEP, que mide la capacidad de los programas para alcanzar sus objetivos, es la variable dependiente (*véase apartado 2.1*) y el porcentaje de ASM atendido de acuerdo a los informes de seguimiento de Sedesol es la variable independiente principal.

También incorporé dos variables de control al modelo: *año de inicio de la evaluación* y *antigüedad del programa*. Por un lado, *año de inicio de la evaluación* evita que los resultados del análisis sean distorsionados por el descenso generalizado y sostenido en la utilización de las evaluaciones que observé en los programas de Sedesol durante el periodo (*véase capítulo 2*). Por otro lado, *antigüedad del programa* permite conocer cuánto tiempo tienen en operación los programas. A partir de este dato, asumo que a mayor antigüedad también es mayor la resistencia de los operadores de los programas a implementar las recomendaciones y cambiar su *modus operandi*.

Por otra parte, consciente de que la evaluación no es lo único que influye en el desempeño de los programas, opté por considerar el papel que juega su implementación, entendida como el proceso que ocurre entre las declaraciones formales de una política y el resultado final alcanzado (Revuelta, 2007). El desarrollo teórico sobre los problemas de la implementación es más antiguo y rico que el estudio sobre el efecto de las evaluaciones, por ello, tomé en consideración dos trabajos destacados el de Luis Aguilar (1993), y el de Weimer y Vining (1992; 2004).

Aguilar (1993) se basa en la teoría de la organización para señalar que los problemas de la implementación son generados por el enfoque *top-down* y también que, en parte, su solución se encuentra en prestar atención al contexto local (enfoque *bottom-*

up).⁶⁷ Weimer y Vining (1992; 2004) identifican tres factores determinantes para la implementación: la lógica de la política (la autoridad legal otorgada para obtener los resultados deseados), la coordinación requerida (la participación de quienes controlan los elementos necesarios), y las habilidades y compromiso de los gestores (los motivaciones y recursos de los gestores).

De las investigaciones mencionadas, retomé elementos de la propuesta de Weimer y Vining (2004) para introducir tres variables que, entre otras, ayudan a entender una exitosa implementación de políticas públicas: el presupuesto anual del programa, la temática social atendida y el tipo de dependencia responsable de su operación. La primera, refleja el tamaño o importancia del programa valorado, mientras que las otras dos exponen diferencias en su implementación.

Estas variables permitieron que, además del efecto de las evaluaciones, el modelo estadístico aplicado considerara una explicación alternativa para la valoración del desempeño de los programas (INDEP). Cabe aclarar que decidí excluir del análisis las dinámicas intergubernamentales que se generan alrededor de la implementación de los programas, a pesar de su relevancia, debido a las limitaciones de información y de recursos de esta investigación.

A continuación, describo brevemente las seis variables independientes consideradas en el modelo y cuáles son las hipótesis en torno a cada una de ellas y su efecto sobre el desempeño de los programas.

⁶⁷ Aguilar (1993) señala que el enfoque gerencial de arriba hacia abajo, que se caracteriza por una relación vertical donde las decisiones se toman en los ámbitos superiores, propicia diversos problemas como: la intervención de muchos poderes durante el proceso; la desorganización de las organizaciones y el comportamiento irresponsable de los operadores, que se manifiesta en la evasión de tareas, resistencia a reglamentaciones y bajo desempeño. Para solucionar estos problemas, Aguilar sugiere dos rutas. Por un lado, la creación de controles mediante reglamentos, personas de confianza, sanciones e incentivos que deben ser establecidos y manejados por un cuerpo directivo enérgico y supervisor, que permita garantizar la fidelidad y la conformidad de los operadores en los términos de la decisión o del programa de gobierno (Aguilar, 1993: 79).

Por otro lado, Aguilar enfatiza la necesidad de prestar atención al contexto local, considerado determinante para el desarrollo y resultado de las políticas al tener una función decisiva de la micro-implementación. Este enfoque asume que “el éxito global de una política depende del desempeño y las consecuencias de la prestación local de servicios, y lo conceptualiza como la ‘adaptación mutua’... que existe entre las condiciones de la agencia local y el programa de acción diseñado en las alturas”. De tal forma que la adaptación se percibe como una prescripción normativa y una variable observable y decisiva en el desarrollo de una política (Aguilar, 1993: 83).

El *porcentaje de ASM atendido* señala qué tanto se utilizan las evaluaciones. En otras palabras, el porcentaje de recomendaciones derivadas de la EED anterior que fueron formalizadas y efectivamente implementadas (t-1). Asumo que entre más ASM sean atendidos mayor es la mejoría en el desempeño de los programas.

El *año de la evaluación*, es una variable de control que modela el descenso gradual observado en la utilización de las evaluaciones entre 2007 y 2014 (*véase capítulo 2*) y, por tanto, evita que se distorsione el análisis. Es decir, introduce al análisis una tendencia generalizada y sostenida que se identifica en los programas de Sedesol diferenciando el momento en que se iniciaron las evaluaciones, tendencia que permite afirmar que los programas emplean cada vez menos sus recomendaciones.

La *antigüedad del programa*, se emplea para representar la resistencia al cambio que podrían presentar las unidades evaluadas debido al tiempo que tienen en operación. Asumo que entre más antiguo sea un programa más fuertes serán las inercias en su implementación y, por tanto, mayor la resistencia al cambio.

El *presupuesto anual del programa* refleja el tamaño e importancia del programa evaluado. Asumo que a mayor presupuesto asignado mayor presión para que el programa atienda las recomendaciones y, por tanto, mayor el impacto de la evaluación. Para suavizar la oscilación de los valores registrados, en el análisis se incluye el logaritmo del presupuesto.

La *temática social atendida por el programa* permite considerar que existen diferentes niveles de complejidad en la operación de los programas valorados, dependiendo de la realidad social que buscan cambiar. Asumo primero, que a mayor complejidad del tema menor es el efecto de las evaluaciones, segundo, que los programas orientados al desarrollo son los más complejos, seguidos de los de género, los productivos, y por último, los de vivienda o espacios públicos.

La sexta y última variable, *tipo de dependencia responsable de su operación*, permite diferenciar entre aquellos programas que son operados solo por Sedesol y en los que participan otras dependencias o entidades. Al respecto asumo que la evaluación tiene menor impacto entre más dependencias o entidades estén vinculadas en la

operación de los programas. Los tipos de dependencias o entidades identificados fueron: Sedesol, órganos desconcentrados, coordinación Nacional o colaboración intersecretarial.

(A) ANÁLISIS DE DATOS

En esta sección, describo los datos empleados en la investigación a partir de la revisión de 75 EED efectuadas a programas de Sedesol, y después presento los resultados del análisis.

Primero, exploro la relación (uno a uno) entre la variable dependiente y las independientes. En el cuadro 4.2, muestro el número de observaciones disponibles, la media, así como los valores mínimos y máximos. Debido a que empleo los ASM atendidos a partir de la EED anterior, el número de observaciones para el análisis se redujo a 47.

El cuadro 4.2 muestra que en promedio los programas obtienen un INDEP de 6.35 (en un rango del cero al diez), que atienden 64 por ciento los ASM y que tienen una antigüedad media de 16 años. El presupuesto de los programas oscila entre 68 y 63,300 millones de pesos, sin embargo, los valores del logaritmo van de 4 a 11. Respecto a las variables *año de la evaluación*, *tema del programa* y *tipo de dependencia(s) responsable(s)*, estas merecen ser analizadas como variables categóricas debido a su naturaleza, por lo cual, no podría hacerse una interpretación correcta de los valores medios observados.

Cuadro 4. 2 Valores medio, mínimo y máximo de las variables

	Observaciones	Máximo	Media	Mínimo	Error estándar
Índice de desempeño (INDEP)	75	10	6.35	1.40	1.65
Aspectos susceptibles de mejoras (t-1)	47	1	0.64	0.07	0.33
Año de la evaluación	75	2014	2010	2008	2.15
Antigüedad	75	65	15.79	0	15.81
Log presupuesto anual del programa	75	11.06	7.09	4.22	1.65
Tema del programa	75	4	3.08	1	1.12
Tipo dependencia(s) responsable(s)	72	3	2	1	0.66

Fuente: Elaboración propia.

En el cuadro 4.3, presento un análisis de correlación entre las variables explicativas y la variación del INDEP. En general, se aprecia una relación de moderada a débil. En este caso, las correlaciones más fuertes registradas respecto a la variable dependiente fueron: *antigüedad del programa* y *tipo de dependencia responsable* de su operación (ambas con 0.29). Además, se observa una relación negativa entre el INDEP y *año en que se inicia la evaluación* (-0.23). Relación prevista a partir de los resultados del capítulo 2, donde se observa que cada vez se aprovechan menos las evaluaciones durante el periodo.

En la matriz de correlación también identifiqué tres casos donde variables explicativas presentaron, entre sí, una correlación de moderada a alta pero sin acercarse de manera preocupante a uno. Por ello, consideré que los resultados no muestran un problema de multicolinealidad que demande cambios en la selección de variables. Sin embargo, a continuación, presento brevemente algunas explicaciones que podría ayudar a entender las correlaciones observadas:

- *La antigüedad del programa y el tipo de dependencia responsable de su operación* (0.33). Esta correlación podría explicarse, por un lado, por la decisión que tomé de no implementar un análisis de panel y de emplear como unidad de medida las evaluaciones en vez de los programas, provocando que el nombre de los programas apareciera tantas veces como fueron valorados y así también la dependencia responsable de su operación. Por otra parte, esta correlación podría entenderse al saber que la mayoría de los programas que no son operados directa o exclusivamente por Sedesol son más antiguos.
- *El logaritmo del presupuesto y la temática del programa* (0.47). Al respecto, durante el análisis de la información, pude observar que los programas orientados al desarrollo reciben, en su mayoría, mucho más dinero que el resto de los programas de Sedesol.
- *El porcentaje de ASM atendido y el año de la evaluación* (-0.60). Como ya he mencionado previamente, esta relación ya estaba prevista desde los hallazgos del capítulo 2, donde identifiqué que en ese periodo y al pasar de los años, los programas utilizan cada vez menos las evaluaciones. De tal forma que entre más reciente sean las evaluaciones se espera encontrar un menor porcentaje de ASM atendido.

Cuadro 4. 3 Matriz de correlación de las variables principales

							Número de Observaciones
							45
	Índice Desempeño	ASM Atendido t-1	Logaritmo Presupuesto	Antigüedad	Tipo Dependencia	Tema Programa	Año Evaluación
Índice de desempeño (INDEP)	1						
Aspectos susceptibles de mejoras (t-1)	0.1604	1					
Log Presupuesto Anual del Programa	0.2379	0.1370	1				
Antigüedad	0.2955	-0.0534	-0.2658	1			
Tema del Programa	0.0930	0.2215	0.4705	0.0767	1		
Tipo Dependencia(s) Responsable(s)	0.2942	-0.0442	0.0576	0.3295	-0.0822	1	
Año de la Evaluación	-0.2335	-0.5970	0.0721	0.0975	0.0851	0.1162	1

Fuente: Elaboración propia.

A partir de los resultados obtenidos en esta primera aproximación, resalta el hecho de que la mayoría de las variables explicativas muestran una relación con la variable dependiente, aunque ésta sea moderada. Sin embargo, es importante tener en cuenta que los valores del cuadro 4.3 son solo correlaciones por pares y no enunciados causales determinantes. Al no controlarse la relevancia de las variables que compiten, se pueden enmascarar relaciones causales más complicadas, como el tamaño y la significancia estadística.

Debido a que hay limitaciones en la matriz de correlación, presento a continuación, un análisis bivariado de MCO con el propósito de conocer la capacidad de cada variable independiente para explicar la estimación del INDEP. Los resultados se presentan en el cuadro 4.4.

En los resultados de la regresión bivariada, noté que cinco de las seis variables independientes arrojaron coeficientes significativos, por lo menos en uno de sus valores cuando se trató de variables categóricas. Sin embargo, *el porcentaje de ASM atendido* la variable de mayor importancia para este análisis no mostró capacidad explicativa.

La ausencia de relación entre el *porcentaje de ASM atendido* y el INDEP se atribuye a la existencia de una tercera variable que suprime la relación entre las dos primeras, en este caso, el *año de la evaluación*. Por tanto, la introduje como variable de control para el análisis de regresión multivariada que presento a continuación.

Cuadro 4. 4 Análisis de regresión bivariada respecto a la valoración del INDEP

	Coefficiente	Prob>F	R^2	Adj. R^2	Núm. Casos
AASM_t-1	.4220586	0.5844	0.0067	-0.0154	47
Log Presupuesto Anual del Programa	.2641818	0.0216**	0.0702	0.0575	75
Antigüedad	.0347248	0.0034***	0.1113	0.0992	75
i. Temática del Programa		0.6687	0.0216	-0.0198	75
Cons (Vivienda o Espacios Públicos)	6.13625***				
Productivo	0.1127501				
Género	-0.4802499				
Desarrollo	0.3927976				
i. Tipo Dependencia		0.0200**	0.1285	0.0917	75
Cons (Sedesol)	5.836***				
Desconcentrada	0.9256667**				
Coordinación Nacional	1.691143**	0.0054***	0.1402	0.1153	72
i. Año de la Evaluación		0.0884*	0.1077	0.0567	75
Cons (2008)	5.574167***				
2009	.5069445				
2010	1.603333**				
2012	1.113526*				
2014	.5783333				

Fuente: Elaboración propia.

(B) RESULTADOS: REGRESIÓN MULTIVARIADA DE MÍNIMOS CUADRADOS ORDINARIOS

Para verificar los supuestos establecidos en torno a la capacidad explicativa de las seis variables independientes seleccionadas, presento tres modelos de regresión multivariada aplicando el método de MCO (véase cuadro 4.5).

En el primero, analizo la relación entre la implementación de los ASM y la estimación del INDEP, la capacidad de los programas para alcanzar sus objetivos. Para ello, incorporo dos variables de control (el *año en que inició la evaluación* y la *antigüedad del programa*), y también otras tres variables que proveen una explicación alternativa a la valoración del INDEP, a través de las cuales rescato diferencias en la implementación de los programas observados (el *presupuesto*, el *tema del programa* y la *entidad responsable de su operación*). Este modelo mostró una capacidad explicativa considerable (R^2 ajustada de 25.7 por ciento), pero la variable independiente de mayor interés no mostró significancia estadística y, además, solo dos de las seis variables fueron significativas (el *presupuesto* y la *antigüedad del programa*). Estos resultados

me permiten concluir que la capacidad explicativa del modelo refiere a la disparidad generada por las otras cuatro variables que no resultaron relevantes. En este sentido, a pesar de su aparente capacidad explicativa y su elevada confiabilidad, este modelo no aporta información valiosa para el análisis.

Para el segundo modelo, opté por explorar únicamente la capacidad explicativa del *porcentaje de ASM atendidos por los programas*, la variable independiente de mayor interés para este análisis, y la variable de control *año de la evaluación*, debido a que esta última resulta necesaria para explorar la relevancia de la primera. Con estas dos variables, la capacidad explicativa del modelo descendió hasta 6.5 por ciento (R^2 ajustada) y la confiabilidad del modelo bajó a 85 por ciento (Prueba de F).

En el tercer modelo, incorporé al análisis anterior una segunda variable de control, la *antigüedad del programa*. Con la variable adicionada, tanto la confiabilidad del modelo como su capacidad explicativa mejoraron para alcanzar una R^2 ajustada de casi 13 por ciento. En este caso, a diferencia del primer modelo, las tres variables analizadas mostraron significancia estadística, todas con una relación positiva con la valoración del INDEP.

De los resultados del último modelo pude concluir lo siguiente: cuando los responsables de los programas implementaron los ASM, la valoración del INDEP aumentó casi dos unidades; los programas evaluados en 2009 mejoraron su desempeño en casi cuatro unidades respecto al 2008, y los evaluados en 2010 solo en dos unidades respecto al 2009 y, por cada año adicional de antigüedad, el INDEP de los programas fue apenas 0.03 puntos mayor.

¿La implementación de los ASM ayuda a explicar los índices de desempeño obtenidos por los programas?

Los resultados del tercer modelo confirman que la implementación de los ASM ayuda a explicar el desempeño de los programas, aunque en una proporción modesta. Y si bien fue necesario introducir la variable de control *año de inicio de la evaluación* para que la variable explicativa principal mostrara significancia estadística, fue un hallazgo

relevante corroborar que la utilización de las evaluaciones tiene un efecto en la gestión pública. Esta afirmación resulta más evidente en los primeros años que se realizaron las evaluaciones.

Cuadro 4. 5 El efecto de las eed en la valoración del INDEP

	Núm Obs 45			Núm Obs 47			Núm Obs 47		
Variación INDEP	Modelo1 (INDEP)			Modelo2 (INDEP)			Modelo3 (INDEP)		
	Coef.	Std Err.	P> t	Coef.	Std Err.	P> t	Coef.	Std Err.	P> t
AASM_t-1	0.1790	1.2062	0.883	2.0369	1.1639	0.087*	1.9309	1.1251	0.094*
i. Año de la Evaluación									
Cons (2009)									
2010	0.3319	0.6939	0.636	1.7295	0.7723	0.03**	1.7451	0.7458	0.024**
2012	-0.9204	0.6692	0.178	-0.0733	0.7746	0.9250	-0.0008	0.7488	0.9990
2014	-1.0123	1.0748	0.353	1.5161	0.9988	0.1370	1.3459	0.9682	0.1720
Antigüedad	0.0674	0.0217	0.004***				0.0303	0.0151	0.051*
Log Presupuesto Anual del Programa	0.3820	0.1692	0.031**						
i. Tipo Dependencia									
Cons (Sedesol)									
Desconcentrada	-0.8271	0.7226	0.261						
Coordinación Nacional	0.9881	0.8018	0.226						
i. Tema									
Cons (Vivienda o Espacios Públicos)									
Productivo	-0.7647	0.8451	0.372						
Género	1.1495	1.1157	0.310						
Desarrollo	-1.3617	0.9238	0.150						
Constante	4.0973	1.6315	0.017	4.1979	1.2205	0.0010	3.7805	1.1966	0.0030
		Prob > F	0.0266			Prob > F	0.1471		
		R²	0.4428			R²	0.1463		
		Adj R²	0.2571			Adj R²	0.0649		
		Prob > F	0.0574			Prob > F	0.0574		
		R²	0.2230			R²	0.2230		
		Adj R²	0.1282			Adj R²	0.1282		

Fuente: Elaboración propia.

Nota: * p < 0.1; **p < 0.05; *** p < 0.01.

REFLEXIONES FINALES

Los análisis presentados en este capítulo permiten observar el efecto (intencionado, inmediato y no inmediato) de las evaluaciones a los programas de Sedesol entre 2007 y 2014. Los hallazgos señalan que las evaluaciones permiten que los programas mejoren tanto su diseño como su desempeño. Pero también, que los beneficios generados en estos dos ámbitos fueron modestos.

En materia de diseño. El efecto inmediato de las evaluaciones, observado a través de los cambios realizados a las Reglas de Operación (ROP), me permitió determinar que las Evaluaciones de Diseño (ED) ayudaron mejorar la lógica interna de 44 por ciento de los 21 programas analizados. En los cuales, a su vez, se implementó 37 por ciento del total de recomendaciones de diseño emitidas que eran aplicables a ROP. En este sentido, advertí un efecto mayor por parte de las evaluaciones realizadas en 2007 y, en especial, sobre las variables más importantes para la construcción de la MIR (indicadores, objetivo del programa y población objetivo).

Los resultados muestran un efecto considerablemente menor en 2011 con relación a 2007. Al respecto, cabe suponer que, en la medida en que la primera evaluación ayudó a los programas a mejorar su lógica interna, éstos requirieron realizar menos ajustes en su diseño.

En un esfuerzo por saber si las recomendaciones atendidas fueron suficientes o adecuadas para mejorar la calidad del diseño de los programas valorados, llevé a cabo un análisis exploratorio donde comparé el número de temas modificados en ROP por programa y la variación en su ICADI. A partir de este análisis, solo me fue posible determinar que se observa una sutil relación entre ambos temas; sin embargo, debido al reducido número de observaciones disponibles en este trabajo no fue posible realizar un análisis estadístico ni determinar qué tan fuerte es esta relación. Por tanto, valdría la pena explorar el alcance de las recomendaciones de diseño en trabajos posteriores.

Considero relevante señalar, además, que los programas estuvieron sujetos a numerosos ajustes no vinculados con las recomendaciones de diseño que emitieron los

evaluadores: en concreto, 36 por ciento de los cambios en ROP observados. En este sentido, considero que también valdría la pena estudiar este fenómeno en una investigación posterior.

En materia de desempeño. El efecto no inmediato de las evaluaciones se estimó a partir de un análisis de regresión lineal multivariada que permitió conocer qué tanto los ASM atendidos ayudan a explicar el desempeño que tuvieron los programas. Este enfoque trasciende la aplicación instrumental de recomendaciones o ASM y se interesa por los beneficios que genera esta práctica en la gestión pública y en el logro de sus objetivos.

El análisis estadístico que realicé en este capítulo, en promedio se atendió 65 por ciento de los ASM derivados de las EED (*véase capítulo 2*), revela que la utilización de las evaluaciones ayuda a explicar entre 7 y 13 por ciento de la valoración del desempeño de los programas, con una confiabilidad del 95 por ciento. Esto significa que, a pesar de la complejidad y la diversidad de factores que influyen en el desempeño de los programas sociales, el efecto de las evaluaciones es modesto pero perceptible. Este hallazgo permite por fin dimensionar, aunque sea sólo para el caso de los programas de Sedesol, el alcance real que tienen los procesos de retroalimentación y mejora continua que se llevan a cabo.⁶⁸

⁶⁸ Por otro lado, cabe señalar que si bien la capacidad explicativa del modelo es baja, se debe tomar en cuenta que el análisis tuvo variables con diferentes unidades de medida y aun así el modelo mostró relaciones significativas.

CAPÍTULO 5

LA UTILIZACIÓN Y EL EFECTO DE LAS EVALUACIONES: ESTUDIO DE CASO DE PDZP (2010 Y 2014) Y 3X1 (2008 Y 2014)

CAPÍTULO 5. LA UTILIZACIÓN Y EL EFECTO DE LAS EVALUACIONES: ESTUDIO DE CASO DE PDZP (2010 Y 2014) Y 3X1 (2008 Y 2014)

En este último capítulo analizo cuál ha sido la utilización de las evaluaciones y qué efecto ha tenido en los programas, con el propósito de ayudar a entender mejor los hallazgos encontrados a lo largo de la tesis. Para ello, utilizaré la información recopilada de las evaluaciones que se hicieron a dos programas orientados a promover el desarrollo: el Programa para el Desarrollo de Zonas Prioritarias (PDZP) y Programa 3X1 para Migrantes (3X1). Ambos eran operados por la Unidad de Microrregiones de Sedesol.

Por un lado, el PDZP, atendía de manera integral los rezagos de las denominadas Zonas de Atención Prioritarias (ZAP). Por otro lado, el 3X1, apoyaba las iniciativas de los migrantes organizados para realizar proyectos que contribuían al desarrollo de sus localidades de origen.

El estudio lo realicé a partir de los hallazgos encontrados en capítulos previos y de entrevistas a evaluadores o funcionarios que estuvieron vinculados al proceso evaluativo de estos dos programas.

El capítulo se compone de tres apartados. En el primero explico brevemente la selección de casos y las características generales de los programas en cuestión. En el segundo, abordo los hallazgos sobre la formalización e implementación de los Aspectos Susceptibles de Mejora (ASM). Por último, en el tercer apartado, exploro los efectos observados de las evaluaciones a partir de cambios en el diseño (cambios en las ROP y en el ICADI) y en el desempeño de los programas (variaciones en el INDEP).

5.1. ESTUDIOS DE CASO SELECCIONADOS

Como mencioné previamente, abordé dos programas de Sedesol orientados a combatir la pobreza a través de la promoción del desarrollo. Por un lado, el PDZP, atendía de manera integral los rezagos de las denominadas Zonas de Atención Prioritarias (ZAP), las cuales se conformaban principalmente por municipios y localidades de muy alta y alta marginación. Éstas últimas ubicadas en municipios de marginación media, baja y muy baja. Por otro lado, el 3X1, se encargaba de canalizar y potencializar las remesas que envían los emigrantes organizados.

Los dos programas seleccionados contaban con un número de evaluaciones igual o superior a la media, entre ellas valoraciones de diseño y de desempeño. Partiendo del hecho de que todos los programas analizados en esta tesis tuvieron un decrecimiento en la atención de ASM atendidos durante el periodo (*véase capítulo 2*), en este capítulo analizo uno que registró un descenso moderado, el PDZP con seis por ciento, y otro con un decrecimiento más importante, 3X1 con 63 por ciento (*véase cuadro 5.3*).

Elegí los años analizados siguiendo dos criterios: por un lado, que se observara algún contraste en la combinación de factores endógenos o exógenos que caracterizaron cada evaluación. Por otro, el mayor distanciamiento temporal posible, por si acaso existieran algún cambio significativo entre las primeras y las últimas evaluaciones realizadas durante el periodo.

Para el caso del PDZP, elegí los años 2010 y 2014. En 2010, los factores endógenos de esta evaluación mostraban valores por debajo de la media del programa, mientras que en 2014 dichos factores se ubicaron por encima. Así, mientras que la evaluación del 2014 tenía mayor difusión en medios electrónicos y hubo una mayor cultura de la evaluación ese año, en 2010 la independencia del proceso fue mayor, ya que éste fue financiado por Coneval. A partir del análisis de factores, consideré que las dos evaluaciones contaban con buenas expectativas para su utilización, pero en la práctica el porcentaje de ASM atendido varió mucho entre ambas, 100 por ciento en 2010 y 28.8 en 2014 (*véase cuadro 5.1 y 5.4*).

Cuadro 5. 1 Evaluaciones realizadas al PDZP

Año	GESOC	Resultados Google	Resultados Google Scholar	Número Evaluaciones	Origen Presupuesto	Término de Contratación	Divulgación
2011	ICADI	149	10	2	Sedesol	Adjudicación directa	Sedesol y Coneval
2009	INDEP	116	6	0	Coneval	Adjudicación directa	Coneval y link Sedesol
2010	INDEP	160	7	1	Coneval	Convenio	Coneval y link Sedesol
2012	INDEP	47	12	3	Sedesol	Convenio	Coneval y link Sedesol
2014	INDEP	372	14	4	Sedesol	Convenio	Coneval
		Media	149	10	2		
		Mínimo	47	6	0		
		Máximo	372	14	4		

Fuente: Elaboración propia.

* Para el caso de Sedesol, la unidad encargada de supervisar todas las evaluaciones es la Dirección General de Evaluación y Monitoreo de los Programas Sociales.

Para el análisis del programa 3x1, elegí los años 2008 y 2014. En 2008, los factores endógenos se encontraban por debajo de la media y en 2014 por encima. Sin embargo, no hay claridad sobre cuál contaba con mayor distancia entre evaluador y evaluado debido a que en 2008 la evaluación la financió Coneval y, mientras lo hizo, la contratación fue directa. En el caso de 2014, el ajuste lo llevó a cabo Sedesol a través de un convenio entre esta Secretaría y el COLMEX. El análisis de factores señalaba que la evaluación de 2008 sería menos favorable para su utilización, sin embargo, alcanzó un porcentaje de ASM atendido elevado (92.2 por ciento). Por otro lado, la combinación de factores para 2014, que valoré más idónea, reportó un nivel de cumplimiento muy bajo (25.9 por ciento) (véase cuadro 5.2 y 5.5).

Cuadro 5. 2 Evaluaciones realizadas al Programa 3X1


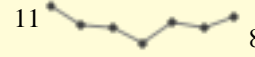






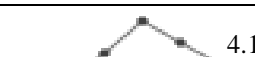
Año	GESOC	Resultados Google	Resultados Google Scholar	Número Evaluaciones	Origen Presupuesto	Término de Contratación	Divulgación
2007	ICADI	31	7	5	Sedesol	Adjudicación directa	Sedesol y Coneval
2011	ICADI	127	25	10	Sedesol	Adjudicación directa	Sedesol y Coneval
2008	INDEP	45	8	7	Coneval	Adjudicación directa	Coneval y link Sedesol
2009	INDEP	115	15	8	Coneval	Adjudicación directa	Coneval y link Sedesol
2010	INDEP	134	13	9	Coneval	Convenio	Coneval y link Sedesol
2012	INDEP	42	13	11	Sedesol	Convenio	Coneval y link Sedesol
2014	INDEP	349	18	12	Sedesol	Convenio	Coneval
		Media	115	13	9		
		Mínimo	31	7	5		
		Máximo	349	25	12		

Fuente: Elaboración propia.

* Para el caso de Sedesol, la unidad encargada de supervisar todas las evaluaciones es la Dirección General de Evaluación y Monitoreo de los Programas Sociales.

En el cuadro 5.3 presento una síntesis de los hallazgos encontrados en torno a la utilización y el efecto de las evaluaciones de los dos programas mencionados. Hallazgos que son explorados a lo largo de este capítulo con la intención de alcanzar un nivel mayor de comprensión sobre lo observado.

Cuadro 5. 3 Síntesis de los hallazgos generales sobre los estudios de caso

	PDZP (2009 - 2014)	3x1 (2007 - 2014)
Núm. total evaluaciones Me = 5	5 (1 ED + 4 EED)	7 (2 ED + 5 EED)
Costo total de las evaluaciones	\$ 953,067.80	\$ 1,496,265.27
Total recomendaciones emitidas	26	43
Max.	8	11
Min.	3	1
Variación recomendaciones	6 	11 
Promedio de recomendaciones formalizadas Me = 58%	53.0%	86.4%
Variación recomendaciones formalizadas	2 	10 
Porcentaje de ASM atendido (promedio) Me = 65.11%	47.3%	63.5%
Variación ASM atendidos	- 6.3%	- 62.8%
	35.0 	88.8 
Promedio de ASM de ED atendidas	57.9%	83.9%
ASM sobre población objetivo atendidos	No	Sí
Promedio de ASM de EED atendidas	44.6%	53.3%
Variación en el ICADI (2007 y 2011)	n.a.	2.85
	2011 => 6.85	3.60 
Promedio de ASM de EED atendido	44.56%	53.29%
Variación en el INDEP (2008, 2009, 2010, 2012 y 2014)	- 2.35	1.57
	6.41 	2.62 

Fuente: Elaboración propia.

Nota: n.a. = no aplica.

5.1.1. PROGRAMA PARA EL DESARROLLO DE ZONAS PRIORITARIAS (PDZP), 2009 – 2014

El PDZP se creó en 2009 a partir de la fusión de tres programas que eran operados por Sedesol (Desarrollo Local, Microrregiones y Apoyo a Zonas de Atención Prioritaria). Éste atendió de manera integral los rezagos vinculados con la infraestructura básica comunitaria y la carencia de servicios básicos en las viviendas, ubicadas en los municipios de muy alta y alta marginación denominados ZAP o en otras localidades, territorios o regiones que también presentaron condiciones de rezago.⁶⁹

Este programa desempeñó cinco acciones orientadas a: i) mejorar la calidad y espacios en la vivienda (muros, techos fijos, pisos firmes y la disminución del hacinamiento); ii) proveer servicios básicos en la vivienda (agua entubada o captación de agua, electricidad, drenaje, baños ecológicos y estufas ecológicas con chimenea); iii) desarrollar infraestructura social comunitaria (redes de distribución de agua, sistemas de recolección de agua, redes de drenaje pluvial y sanitario, plantas de tratamiento de aguas residuales y para consumo humano, redes para la distribución de energía eléctrica, centros comunitarios, albergues, sistemas de comunicación, caminos rurales y puentes, bancos de alimentos, infraestructura productiva); iv) brindar apoyos complementarios para la promoción y organización social comunitaria o levantamiento de información que permita mejorar las inversiones del Programa; y v) apoyar situaciones de emergencia o en caso de contingencia por fenómenos naturales.

Para esta tesis analicé cinco evaluaciones realizadas al PDZP, una de diseño y cuatro de desempeño. Todas fueron realizadas por el COLMEX con un costo total aproximado de 953 mil pesos.

⁶⁹ <http://plataformacelac.org/programa/108>

Cuadro 5. 4 Evaluaciones analizadas del Programa para el Desarrollo de Zonas Prioritarias

Año	Tipo	Evaluador		Coordinador	Costo (Miles)	Presupuesto (MDP)	ASMA
		Organismo	Área				
2011	ED	COLMEX	Centro de Estudios Económicos	Gerardo Esquivel Hernández	556.80	6,194.29	58%
2009	EED	COLMEX	Centro de Estudios Demográficos y Desarrollo Urbano	Martha Rosa Schteingart Garfunkel	82.15	8,160.99	35%
2010	EED	COLMEX	Centro de Estudios Demográficos y Desarrollo Urbano	Martha Rosa Schteingart Garfunkel	74.12	7,084.53	100% ←
2012	EED	COLMEX	Centro de Estudios Demográficos y Desarrollo Urbano	Clara Eugenia Salazar Cruz	100.00	6,411.09	14.5%
2014	EED	COLMEX	Centro de Estudios Demográficos, Urbanos y Ambientales	Luis Jaime Sobrino Figueroa	140.00	6,883.48	28.8% ←

Fuente: Elaboración propia.

Tipos de evaluaciones: ED = Evaluación de Diseño; EED = Evaluación Específica de Desempeño.

5.1.2. PROGRAMA 3x1 PARA MIGRANTES (3x1), 2007 - 2014

Este programa nace en 2002 y apoya las iniciativas de los migrantes organizados para realizar proyectos que contribuyan al desarrollo de sus localidades de origen, mediante la aportación de los tres órdenes de gobierno: federal, estatal y municipal, así como de organizaciones de migrantes en el extranjero.

El programa 3x1 contó con cuatro tipo de apoyos: i) Proyectos de Infraestructura Social. Construir, ampliar o rehabilitar de infraestructura social básica (redes de agua, saneamiento y potabilización, drenaje, alcantarillado y electrificación) o para el mejoramiento urbano y/o protección del medio ambiente (calles, banquetas, zócalos, parques, pavimentos, caminos, carreteras y obras para la conservación de recursos naturales); ii) Proyectos de Servicios Comunitarios (becas académicas o apoyos para el aprendizaje, y espacios de beneficio comunitario); iii) Proyectos Educativos (equipamiento de escuelas o mejoramiento de infraestructura escolar); y iv) Proyectos Productivos (individuales, familiares, o comunitarios).

Cuadro 5. 5 Evaluaciones analizadas del Programa 3x1 para Migrantes

Año	Tipo	Evaluador		Coordinador	Costo (Miles)	Presupuesto (MDP)	ASMA
		Organismo	Área				
2007	ED	CIDE	División Estudios Políticos	Francisco Javier Aparicio Castillo	437.32	220.00	88.8%
2011	ED	COLMEX	Centro de Estudios Económicos	Gerardo Esquivel Hernández	556.80	557.75	79%
2008	EED	COLEF	Departamento de Estudios de Población	René Martín Zenteno Quintero	105.87	503.47	92.2%
2009	EED	COLMEX	Centro de Estudios Demográficos y Desarrollo Urbano	Martha Rosa Schteingart Garfunkel	82.15	562.38	65%
2010	EED	COLMEX	Centro de Estudios Demográficos y Desarrollo Urbano	Martha Rosa Schteingart Garfunkel	74.12	557.47	n.d.
2012	EED	COLMEX	Centro de Estudios Demográficos y Desarrollo Urbano	Martha Rosa Schteingart Garfunkel	100.00	507.75	30%
2014	EED	COLMEX	Centro de Estudios Demográficos, Urbanos y Ambientales	Vicente Ugalde Saldaña	140.00	505.27	26%

Fuente: Elaboración propia.

Tipos de evaluaciones: ED = Evaluación de Diseño; EED = Evaluación Específica de Desempeño.

En este caso, exploré siete evaluaciones: dos de diseño y cinco de desempeño. Al igual que en el programa anterior, la mayoría elaboradas por el COLMEX, solo dos por otros organismos evaluadores, el CIDE en 2007 y el COLEF en 2008. El costo de estas evaluaciones ascendió aproximadamente a un millón 496 mil pesos.

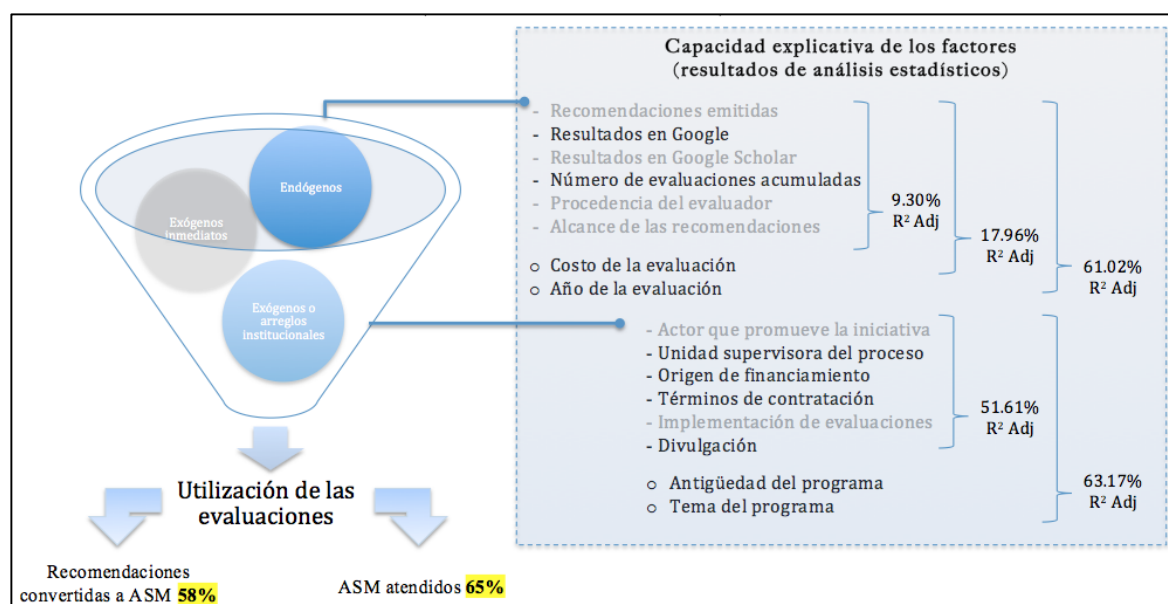
Por otro lado, resalta que en ambos programas el Dr. Esquivel estuvo a cargo de ED, mientras que la Dra. Schteingart coordinó cinco de las nueve EED observadas en los dos estudios de caso. Además, a pesar de la gran diferencia de tamaño de los dos programas, estimado a partir del presupuesto asignado, las ED llegaron a ser igual de costosas.

En los siguientes dos sub-apartados señalo detalles de la utilización y el efecto de las evaluaciones realizadas a los estudios de caso.

5.2. LA UTILIZACIÓN DE LAS EVALUACIONES

Los hallazgos encontrados en capítulos anteriores señalan que, en promedio, el 58 por ciento de las recomendaciones emitidas por los evaluadores son formalizadas como ASM y que, de ese porcentaje, solo el 65 por ciento es implementado. Asimismo, que la difusión de los informes (resultados en Google), la intensidad de las evaluaciones (número de evaluaciones acumuladas), el costo de las mismas, así como la distancia entre evaluador y evaluado son factores que ayudan a explicar parcialmente el que unas sean más utilizadas que otras (*véase ilustración 5.1*).

Ilustración 5. 1 Hallazgos sobre la utilización de la evaluación y la capacidad explicativa de los factores endógenos y exógenos



Fuente: Elaboración propia.

Respecto a la utilización de los programas seleccionados como estudios de caso, cabe señalar que el PDZP, en promedio, formalizó menos recomendaciones y atendió menos ASM que el resto de los programas analizados. Por su parte, el 3X1 superó el promedio de recomendaciones formalizadas y registró un porcentaje de ASM atendidos solo un poco debajo de la media registrada. Estos datos me permiten señalar que el programa 3X1 utilizó más sus evaluaciones respecto al PDZP, a pesar de que recibió mayor número de recomendaciones (*véase cuadro 5.3*).

5.2.1. FORMALIZACIÓN E IMPLEMENTACIÓN DE LOS ASM

En este apartado analizo las recomendaciones emitidas en los informes de evaluación y los ASM atendidos de las cuatro evaluaciones seleccionadas, dos del PDZP y dos del 3X1. Al momento de revisar el cumplimiento de los ASM, es importante tener presente que su seguimiento no se limita a un ciclo presupuestario, sino que los compromisos pueden ser atendidos a lo largo de varios años.

(A) PDZP: ESCENARIOS CON RESULTADOS DISTINTOS 2010 Y 2014

Ambas evaluaciones fueron realizadas por el mismo organismo evaluador (COLMEX), fueron contratadas a partir de un convenio y su difusión en medios electrónicos (resultados en Google) fue superior a la media registrada por el programa. Sin embargo, mientras que la evaluación de 2010 fue financiada por Coneval y sus resultados divulgados tanto en el portal de internet de esta instancia como en la de Sedesol, en la evaluación 2014, cuyo financiamiento estuvo a cargo de Sedesol, solo me fue posible encontrar el informe de esta evaluación en el portal de Coneval debido a los cambios que realizó la actual administración en los portales de las secretarías federales (*véase cuadro 5.1*).

- 2010. En este informe identifiqué cinco recomendaciones, dos señaladas como posibles ASM y otras tres solo mencionadas en el cuerpo del documento. Destaca que las primeras dos también están presentes en la evaluación del año anterior (realizada por la misma coordinadora, la Dra. Schteingart y para las cuales se estableció el año 2012 como fecha límite para su cumplimiento). Para 2011, de las dos recomendaciones repetidas por el mismo evaluador solo se cumplió una: “Documentar una estrategia de cobertura para los Programas con enfoque Territorial”, misma que al inicio de la evaluación en 2010 ya tenía un avance del 20 por ciento y para 2011 alcanza el 100 por ciento (*véase cuadro 5.6*).
- 2014. En este informe identifiqué cuatro recomendaciones, dos reconocidas como posibles ASM y otras dos señaladas en el cuerpo del documento. Al concluir la evaluación, efectivamente se formalizaron los dos ASM previstos, uno específico:

“Elaborar un diagnóstico y un plan de trabajo específico para atender la problemática de los Centros públicos de Cómputo con acceso a internet”, que alcanzó un avance del 50 por ciento para 2011. También un ASM institucional: “Desarrollar una estrategia de difusión... que contenga los elementos y materiales... suficientes que permita garantizar el acceso a la información del Programa a todos los interesados”, del cual, a su vez, se desprendía la elaboración de un documento con la estrategia de difusión a mediano y largo plazo para 2015, y el diseño de los medios de comunicación impresos y electrónicos que se utilizarían para 2016. Sin embargo, a un año de su formalización, el porcentaje de avance registrado para este ASM fue de solo el 7.5 por ciento (*véase cuadro 5.6*).

En general, en estas dos evaluaciones del PDZP se observa que sólo se formalizaron e implementaron ASM recomendados por los evaluadores, estos últimos se encuentran enlistados en todos los informes de las EED.

Cuadro 5. 6 Recomendaciones y ASM derivados de las EED del PDZP, 2010 y 2014

2010		
Recomendaciones emitidas por el evaluador	ASM Sí/No	Avance %
1. Documentar una estrategia de cobertura para los programas con enfoque territorial (<i>ASM 2009 para 2012</i>). - Elaborar un documento que considere un marco teórico, modelos de cobertura, así como una propuesta concreta de estrategia territorial al cual se apega la operación del PDZP.	Sí (Esp.)	100
2. Promover la cultura del mantenimiento, control y vigilancia continua de las obras y acciones a través de la capacitación y fortalecimiento de comités de contraloría social del ejercicio fiscal 2011 (<i>ASM 2009 para 2012</i>).	No	--
* En las ROP 2011, el programa establece entre los criterios de selección dar prioridad a las obras y acciones que incluyan participación financiera de estados y municipios. Se recomienda incluir para la próxima evaluación información que permita conocer el grado de involucramiento de los estados y municipios (<i>aplicable a ROP</i>).	No	--
* Se sugiere la incorporación de documentos que muestren (más allá de las ROP y los <i>Lineamientos de Operación</i>), información sobre la conformación de mecanismos de sensibilización de la población para fomentar la apropiación de obras por parte de los beneficiarios.	No	--
* Aun cuando se reconocen las dificultades de utilizar el <i>Índice de Marginación Absoluta</i> para relacionar el actual indicador de Fin, se reitera la necesidad de buscar otro indicador para confrontarlo de tal forma que permita valorar la incidencia del programa en el cumplimiento del Objetivo 3 del PND.	No	--

2014		
Recomendaciones emitidas por el evaluador	ASM Sí/No	Avance %
1. Elaborar un diagnóstico y un plan de trabajo específicos para atender la problemática de los Centros públicos de cómputo con acceso a internet. - Elaborar un plan de trabajo específico para atender la problemática de los CCA's; - Elaborar un diagnóstico que permita determinar las necesidades técnicas de los CCA's	Sí (Esp.)	50
2. Desarrollar una estrategia de difusión del programa a mediano y largo plazo que contenga los elementos y materiales impresos y electrónicos suficientes que permitan garantizar el acceso a la información del programa a todos los interesados. - Diseñar los medios de comunicación impresos y electrónicos que se utilizarán en la estrategia de difusión del PDZP. - Elaborar un documento que contenga la estrategia de difusión a mediano y largo plazo que permita garantizar el acceso a la información del Programa a todos los interesados.	Sí (Inst.)	7.5
* Se deberá trabajar en el diseño de Indicadores de Resultados que sean más sencillos de leer, que no midan acciones acumuladas, que no sean de comportamiento descendente y que permitan evaluar el desempeño anual del programa.	No	--
* Es aconsejable que el programa establezca acuerdos de correspondencia más asertivos con gobiernos estatales y municipales para el destino de recursos a la operación y mantenimiento de proyectos de infraestructura social y comunitaria (<i>aplicable a ROP</i>).	No	--

Fuente: Elaboración propia con datos de los informes de evaluación y de los *informes de seguimiento de los Aspectos Susceptibles de Mejora publicados por Sedesol publicados entre 2007 y 2014*.

* Recomendaciones identificadas en los informes de evaluación que no formaban parte del listado de posibles Aspectos Susceptibles de Mejora.

Nota: Esp. = ASM Específicos, Inst. = ASM Institucionales.

(B) 3X1: ESCENARIOS QUE CONTRADICEN LOS SUPUESTOS 2008 Y 2014

Si bien las dos evaluaciones observadas fueron realizadas por organismos públicos (COLEF en 2008 y COLMEX en 2014), la primera fue contratada por *adjudicación directa* mientras que la segunda a partir de un *convenio*. En 2008, la difusión de los resultados en medios electrónicos fue muy baja, ubicándose por debajo de la media (45 resultados en Google), la evaluación fue financiada por Coneval y el informe se divulgó tanto en el portal de internet de esta instancia como en el de Sedesol.

Por otro lado, la difusión de la evaluación en 2014 fue mucho más alta (349 resultados en Google), su financiamiento provino de Sedesol y su informe solo estuvo accesible en el portal de Coneval debido a los recientes cambios realizados en los sitios web de internet de las secretarías federales (*véase cuadro 5.8*).

- 2008. En este informe identifiqué cinco recomendaciones señaladas como posibles ASM. Sin embargo, todas estaban vinculadas a las recomendaciones de la ED anterior. Además, entre los cinco ASM que efectivamente fueron formalizados solo uno, de carácter específico, tenía vínculo con el informe: Mejorar los sistemas de información que permitan la descripción y localización de las obras”; otro institucional, no mencionado en las recomendaciones, también provenía de la evaluación anterior: “Mejorar los mecanismos de verificación de la ejecución y terminación de las obras”. Ambos fueron atendidos casi por completo para 2009.

En el caso de los otros tres ASM específicos formalizados para este año, no encontré antecedentes documentados en la evaluación pero observé que también fueron atendidos casi en su totalidad para 2009 (*véase cuadro 5.6*).

- 2014. En esta evaluación destaca que las siete recomendaciones que fueron plasmadas en el informe como posibles ASM fueron formalizadas. A diferencia de los casos previamente analizados, en éste, los evaluadores retomaron de manera explícita las recomendaciones emitidas. Sin embargo, a pesar de que todos los ASM fueron comprometidos para el 2015, en el informe de seguimiento se aprecia que cuatro de carácter institucional prácticamente no lograron avance (oscilando entre 0 y el 7 por ciento). Por otra parte, en cuanto a los ASM específicos, dos alcanzaron un avances modesto (20 o 30 por ciento) y solo uno se atendió en su totalidad: “incorporar en la página de internet del programa una sección que permita a los migrantes orientar sus proyectos a abatir las principales carencias de sus localidades, así como información para entender mejor los conceptos de carencias sociales” (*véase cuadro 5.6*).

De estas dos evaluaciones del programa 3X1 que analicé, llama la atención que la primera, correspondiente al año 2008, retomó elementos de la valoración anterior a pesar de que eran de diferente naturaleza (ED VS. EED) y de haber sido realizada por instituciones distintas (COLEF VS. COLMEX) (*véase cuadro 5.3*). En ésta, se insistió en mejoras señaladas el año anterior pero los responsables del Programa 3X1 no las retomaron todas y optaron por introducir otras no contempladas en el informe. Por otra parte, destaca que en la evaluación de 2014, las recomendaciones, a pesar que se formalizaron en su totalidad, tuvieron un porcentaje de cumplimiento muy bajo.

Cuadro 5. 7 Recomendaciones y ASM derivados de las EED del Programa 3x1, 2008 y 2014

2008		
Recomendaciones emitidas por el evaluador	ASM Sí/No	Avance %
1. Revisar los mecanismos de selección de proyectos y localidades atendidas por el programa (<i>ASM 2007 para 2008, aplicable a ROP</i>).	No	--
2. Revisar la focalización del programa en localidades de alta y muy alta marginación (<i>ASM 2007 para 2008, aplicable a ROP</i>).	No	--
3. Ampliar la incidencia de las representaciones en Estados Unidos (esquema de representantes, firma de convenios y medir incidencia) (<i>ASM 2007 para 2010</i>).	No	--
4. <u>Recomendación:</u> Especificar la información que los clubes de migrantes que participan en el Programa deberán compartir con el Programa (<i>ASM 2007 para 2009, aplicable a ROP</i>). <u>ASM:</u> Mejorar los sistemas de información que permitan la descripción y localización de las obras. - Definir los elementos mínimos que deberán contener las descripciones de las obras realizadas por los programas a cargo de Sedesol para que estas sean incorporadas al Sistema de Información.	Sí (Esp.)	95
Aspecto 5. Fortalecer la capacitación a municipios y delegaciones estatales en el uso de los sistemas de información en los que se reportan los avances del Programa (<i>ASM 2007 para 2009</i>).	No	--
ASM no vinculados con los recomendados emitidas por los evaluadores	ASM Sí/No	Avance %
1. Continuar con el proceso de mejora de las Matrices de Indicadores para Resultados (MIR), documentando la solidez técnica de los indicadores, su relevancia, el establecimiento de líneas de base y metas. - Fomentar las mejoras en el diseño de indicadores del ejercicio fiscal correspondiente a través de: a) Comentarios y recomendaciones a las fichas técnicas de la MIR, b) Documentación de aspectos relacionados a la calidad de los indicadores en notas o fichas complementarias con base en información disponible. - Fomentar que en las revisiones y comentarios a las MIR, que realicen las instancias correspondientes (SHCP, SFP, Coneval, Inmujeres, ASF, etc.) se considere la documentación técnica disponible, a fin de que las propuestas de mejora sean factibles técnica y económicamente. - Definir, en coordinación con las áreas involucradas, el proceso interno de mejora de MIR del ejercicio fiscal 2011, a fin de que se lleve a cabo de manera ordenada y conforme a la normatividad.	Sí (Inst.)	100 0 100
2. Actualizar y difundir los diagnósticos de los problemas que atienden los programas sociales, documentando la medición de la población potencial y objetivo para su incorporación en las estrategias de cobertura, cambios en el diseño y definición de estrategias óptimas de atención. - Difusión del Diagnóstico. - Actualización de las poblaciones potencial y objetivo.	Sí (Inst.)	100 90
3. Contar con una agenda de evaluación de resultados de los programas sociales, privilegiando las evaluaciones de impacto y de seguimiento a beneficiarios, previo análisis de factibilidad. - Integración de una Agenda de Evaluación 2010 (Avance 100%); - Agenda de evaluación del Programa 3x1 para Migrantes; - Proponer la creación del Comité interno de planeación estratégica, evaluación y mejora de los programas sociales; - Definir la agenda anual de estudios y evaluaciones 2011, con base en el PAE.	Sí (Inst.)	100 25 100 100

4. Mejorar los mecanismos de verificación de la ejecución y terminación de las obras (<i>ASM 2007 para 2008</i>). - Establecer los procedimientos de gestión que promuevan la verificación de la ejecución y terminación de las obras.	Sí (Esp.)	100
2014		
Recomendaciones emitidas por el evaluador	ASM Sí/No	Avance %
1. Realizar un documento que contenga una estrategia de comunicación en el corto y mediano plazo. - Elaborar los contenidos de los instrumentos de comunicación, por ejemplo, folletos, banners para la página en internet, etc.; - Realizar un documento que contenga una estrategia de comunicación en el corto y mediano plazo.	Sí (Inst.)	7.5
2. <u>Recomendación:</u> Establecer en colaboración con el IME en las ROP los criterios y procedimientos de Toma de Nota que faciliten la validación en el registro de los clubes de migrantes y sus integrantes (<i>aplicable a ROP</i>). <u>ASM:</u> Homogeneizar en ROP el procedimiento por el que se valida a los integrantes de los clubes de migrantes en el proceso de Toma de Nota.	Sí (Inst.)	0
3. Incorporar en las ROP 2016 del Programa un párrafo en el que se mencione que los migrantes pueden solicitar información referente a sus proyectos, así como los argumentos por los que éstos fueron apoyados o rechazados (<i>aplicable a ROP</i>).	Sí (Inst.)	0
4. Establecer en las ROP 2016 del programa que se elimina el reembolso de recursos federales en los proyectos productivos, otorgándolos como subsidio completo (<i>aplicable a ROP</i>).	Sí (Inst.)	0
5. Documentar el procedimiento para la determinación de metas del programa 3x1. - Documentar el procedimiento interno que utiliza el 3x1 para la determinación de metas; - Generar un documento que contenga la metodología que se utiliza para la determinación de metas del Programa.	Sí (Esp.)	30
6. Incorporar en la página de internet del Programa una sección que permita a los migrantes orientar sus proyectos a abatir las principales carencias de sus localidades, así como información para entender mejor los conceptos de carencias sociales.	Sí (Esp.)	100
7. Agilizar la implementación del SIG-P3x1, que permitirá a los clubes de migrantes y a sus miembros, dar seguimiento puntual a sus propuestas de proyectos a todo lo largo del proceso, desde la recepción, la validación, la dictaminación, la ejecución y la entrega de obras o apoyos.	Sí (Esp.)	20

Fuente: Elaboración propia con datos de los informes de evaluación y de los *informes de seguimiento de los Aspectos Susceptibles de Mejora publicados por Sedesol publicados entre 2007 y 2014*.

* Recomendaciones identificadas en los informes de evaluación que no formaban parte del listado de posibles Aspectos Susceptibles de Mejora.

Nota: Esp. = ASM Específicos, Inst. = ASM Institucionales.

(C) CONCLUSIONES SOBRE LA FORMALIZACIÓN E IMPLEMENTACIÓN

En general, a partir de la investigación realizada y, en concreto, del análisis de los estudios de caso, pude apreciar que la formalización de ASM es un proceso más complejo de que lo preví al iniciar esta tesis, pues no radica solo en la aceptación (total o parcial) o el rechazo de las

recomendaciones emitidas, sino que se involucran otros escenarios: ASM no identificados entre las recomendaciones, repetición de ASM señalados por el mismo evaluador que es contratado de manera consecutiva dos o más años para un mismo programa, repetición de ASM señalados por otro evaluador en informes anteriores y la aceptación aparente de recomendaciones que en la práctica no derivan en acciones de mejora.

En el caso de Sedesol, los ASM son definidos a partir de negociaciones internas entre la Unidad de Evaluación y los responsables de los programas, mismos que se encuentran limitados por la disponibilidad de recursos y atribuciones por parte de los últimos. En este sentido, la Unidad de Evaluación desempeña un papel muy importante en el aprovechamiento de los informes al promover la implementación de las recomendaciones desde una posición privilegiada en la Secretaría.

Desde la perspectiva de un evaluador contratado por Coneval, en entrevista, la Dra. Schteingart señaló con base en su experiencia que para ella sería imposible explicar el proceso de formalización de los ASM. Afirmó que su equipo tuvo reuniones de trabajo con los responsables de los programas pero que el tema no fue abordado ni durante ni después de la elaboración de los informes. Por tanto, en su opinión, las reuniones no influyeron directamente en las recomendaciones que emitieron los evaluadores pero sí les permitió alcanzar un mayor conocimiento sobre la realidad del programa y, en algunos casos, tomar en cuenta los retos operativos a los que se enfrentaban.

Por otra parte, desde el enfoque de un evaluador que ha sido contratado tanto por Sedesol como por Coneval, el Dr. Esquivel, en entrevista, señaló que su experiencia como evaluador estuvo determinada, en gran medida, por el actor que contrataba sus servicios. En este sentido, cuando era contratado por Coneval, la interacción con la Secretaría era mucho menor y la libertad al momento de publicar su informe era mayor. Cuando la contratación estaba en manos de Sedesol, el proceso evaluativo implicaba mayor intercambio con los funcionarios de la dependencia. Sin embargo, debido a un rol muy activo que desempeñaba la Unidad de Evaluación, más que un mediador entre el evaluador y la unidad responsable del programa valorado, esta unidad llegaba a fungir como un conciliador que, desde un enfoque pragmático, en algunas ocasiones podía favorecer la formalización de compromisos laxos reduciendo el impacto de las evaluaciones, y en otras ocasiones, llegar a generar presión hacia los evaluadores para modificar sus hallazgos o recomendaciones.

Las dos experiencias previamente mencionadas me permitieron corroborar que, efectivamente, un mayor distanciamiento entre evaluador y evaluado favorece la libertad con la que el primero puede desempeñar sus actividades, mismas que en el caso mexicano, desde un inicio se encuentran restringidas por Términos de Referencia, que si bien ayudan a homogenizar los procesos, también obligan a los evaluadores a plasmar sus resultados en determinados formatos y seguir metodologías específicas que restringen la creatividad del evaluador, los temas a tratar, así como las respuestas y las fuentes de información que serán implementadas (en este caso, datos oficiales que proveen las mismas entidades o dependencias evaluadas).

Por otra parte, el exfuncionario “B”, señaló que si bien la independencia del proceso evaluativo aumentó con las contrataciones realizadas por Coneval, la interacción entre evaluadores y evaluados disminuyó considerablemente, experimentándose “un distanciamiento”. Al respecto, aclaró que antes de la creación de Coneval, la unidad de evaluación tenía un rol muy activo y convocaba a numerosas reuniones de trabajo donde había comunicación directa con los evaluadores, mismas que en algunos casos eran tan numerosas que llegaban a desgastar la relación entre las partes. También especificó que, a partir de la creación del Coneval, la participación de la unidad de evaluación disminuyó considerablemente al igual que el número de reuniones, en algunos casos insuficientes, donde ahora la comunicación se da principalmente con los asistentes de los evaluadores.

Además, desde la experiencia del exfuncionario de Sedesol “B”, los factores más importantes para la utilización de las evaluaciones eran: la voluntad de los jefes de los programas, quienes muchas veces se negaban a adoptar recomendaciones argumentando que ellos conocían mejor que los evaluadores los problemas de su programa o, en su caso, los más receptivos, sugerían alternativas consideradas más viables que las recomendadas en principio; la insistencia por parte de la Dirección General de Evaluación sobre la implementación de las recomendaciones⁷⁰ y, en la unidad de microrregiones, la voluntad de

⁷⁰ En entrevista, el exfuncionario de Sedesol “B”, señaló que antes de que se estableciera el PbR, una manera de presionar a los responsables de los programas para la adopción de ciertas recomendaciones era el proceso de validación de las ROP, en el cual la unidad de planeación tenía la última palabra, sirviendo este proceso como mecanismo útil para la negociación. Además, refirió que con el PbR, este recurso pierde fuerza y la atención se vuelca hacia los ajustes a la MIR.

la DGA encargada de determinar el presupuesto, quien en la mayoría de los casos tenía la última palabra sobre los ajustes en los programas a pesar de que era otra DGA la que participaba en los procesos evaluativos.

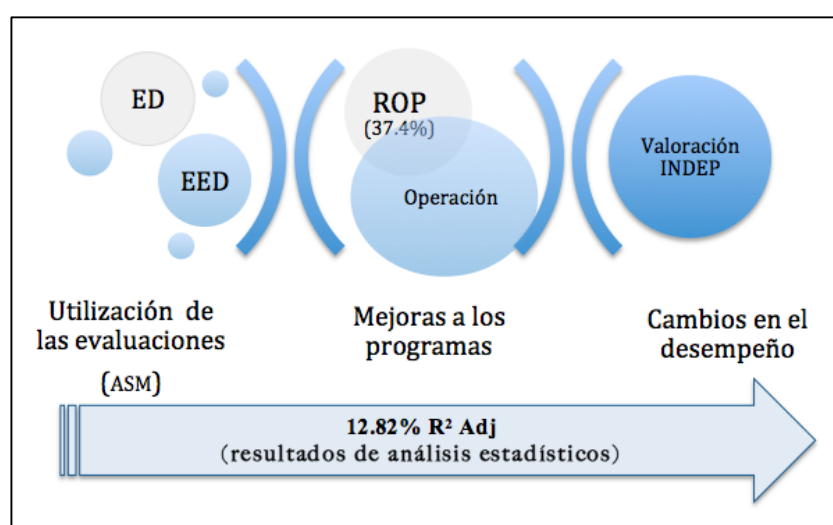
Por último, cabe señalar que la mayoría de los entrevistados coincidieron en que, en la práctica, el incumplimiento de los ASM establecidos no ha derivado en sanciones administrativas. En torno a este hecho, existen posiciones encontradas sobre la conveniencia de establecer recomendaciones vinculativas a fin de imponer un cumplimiento obligatorio. Sin embargo, llama la atención que poco se debate sobre la necesidad de establecer incentivos novedosos para los diferentes usuarios.

5.3. EL EFECTO DE LAS EVALUACIONES EN LOS PROGRAMAS

El mejoramiento de los programas puede tener diferentes explicaciones y manifestarse en diversos ámbitos debido a la complejidad de la APF. Sin embargo, en este apartado abordo específicamente el efecto de la evaluación en el desempeño de los programas valorados.

Para explorar el efecto de la evaluación, retomo los estudios de caso seleccionados al inicio de este capítulo, dos EED de PDZP y dos del programa 3X1. A pesar de que seleccioné solo EED, menciono también los avances de estos dos programas en materia de diseño, ya que asumo que la mejora experimentada en el desempeño también se explica, hasta cierto punto, con los ajustes en su lógica interna (*véase ilustración 5.2*).




Ilustración 5. 2 Hallazgos sobre el efecto de la práctica evaluativa en el desempeño de los programas: la valoración del INDEP y la capacidad explicativa de la implementación de los ASM



Fuente: Elaboración propia.

En el cuadro 5.8 presento los resultados que obtuve en el capítulo 4. De manera sintética, expongo el efecto de la evaluación en el PDZP y en 3X1. En este cuadro, destacan algunos contrastes entre los programas: por un lado, que el PDZP mostró un retroceso en el INDEP; por otro, que el programa 3X1 logró un avance modesto, a la vez que alcanzó un porcentaje más alto de cumplimiento de ASM, tanto de diseño como de desempeño.

Cuadro 5. 8 Síntesis del efecto de la evaluación en los estudios de caso

	PDZP	3x1
Promedio de ASM de ED atendidas Me = 82.00%	58.00%	83.88%
Variación en el ICADI 2007 Y 2011	n.a.	2.85
	2011 => 6.85	 <p>3.60 6.45</p>
Promedio de ASM de EED atendidas Me = 56.27%	44.56%	53.29%
Variación en el INDEP 2008-2014	- 2.35	1.57
	 <p>6.41 4.06</p>	 <p>2.62 4.19</p>

Fuente: Elaboración propia.

Nota: n.a. = no aplica.

A continuación, señalo los cambios que experimentaron los dos programas en su diseño (ROP y el ICADI) y en su desempeño (INDEP y distribución espacial de recursos).

5.3.1. LOS CAMBIOS EN EL DISEÑO

Al explorar los cambios que experimentaron los programas evaluados en su lógica interna y en su consistencia, mismos que se reflejan en las ROP, esperaba que éstos pudieran explicarse a partir de las recomendaciones de las ED realizadas durante el periodo. Sin embargo, de acuerdo con los resultados obtenidos en el apartado anterior y en el capítulo 4 de esta tesis, donde analizo el efecto de las evaluaciones, los cambios experimentados no siempre responden a las recomendaciones derivadas de este tipo de evaluaciones.

Asimismo, al inspeccionar las mejoras que experimentaron los programas evaluados a mediano plazo a través del ICADI, noté que la variación de este índice (efecto no inmediato) y los ajustes en las ROP derivados de las ED (efecto inmediato), aunque guardan cierta relación (*véase capítulo 4*), no siempre están vinculados de manera que pueda hablarse de una causalidad. Esto quedó en evidencia al analizar los cambios en el diseño (ajustes a las ROP y variación del ICADI) que experimentaron los dos programas seleccionados como estudios de caso. Dichos cambios los señalo en los siguientes incisos.⁷¹

(A) EFECTO INMEDIATO DE LAS RECOMENDACIONES DE DISEÑO EN LAS ROP

En Sedesol, durante el periodo analizado, la implementación de ASM derivados de las ED fue de 82 por ciento en promedio. Sin embargo, la proporción de recomendaciones orientadas a mejorar las ROP que se aplicó a partir de este tipo de evaluaciones fue solo del 37 por ciento.

De los dos programas seleccionados como estudio de caso, me llamó la atención que, en términos generales, el programa 3X1 fue el que implementó más ASM, pero a su vez no realizó ningún ajuste en sus ROP a partir de las dos ED que se llevaron a cabo durante el periodo (*véase cuadro 5.9*). El PDZP, por su parte, sí implementó las dos recomendaciones de diseño orientadas a mejorar sus ROP.

⁷¹ El análisis de los cambios de diseño a mediano plazo solo aplica para el Programa 3x1 para Migrantes debido a que el PDZP, a pesar de tener origen en la fusión de tres programas, es considerado de nueva creación y solo cuenta con una valoración del ICADI durante el periodo estudiado.

Cuadro 5. 9 Recomendaciones derivadas de las ED implementadas en las ROP

Programa	Año	Recomendaciones Totales	Recomendaciones para ROP	Recomendaciones ROP no atendidas
3X1	2007	11	3	3
3X1	2011	7	4	4
PDZP	2011	3	2	0

Fuente: Elaboración propia.

Como ya señalé anteriormente, la mejora en el diseño de los programas no se explica por completo a través de las recomendaciones derivadas de las ED atendidas. En este sentido, al analizar los estudios de caso, no me sorprendió encontrar que en las EED los evaluadores también emitieron recomendaciones de diseño aplicables a las ROP.

En las EED del PDZP que analicé (2010 y 2014), identifiqué dos recomendaciones, orientadas a mejorar las ROP, que no fueron formalizadas. Por un lado, en 2010, los evaluadores sugieren incluir para la próxima valoración del programa información que permita conocer el grado de involucramiento de las entidades. Esto, debido a que el programa establece para sus ROP 2011 criterios de selección que dan prioridad a las obras y acciones que incluyen participación financiera de estados y municipios. Posteriormente, en esta misma línea, en la evaluación de 2014, los evaluadores aconsejan al programa solicitar acuerdos de corresponsabilidad con gobiernos estatales y municipales para el destino de recursos a la operación y mantenimiento de proyectos de infraestructura.

En las EED del programa 3X1 (2008 y 2014) encontré seis recomendaciones aplicables a ROP. En la evaluación de 2008 identifiqué tres que daban seguimiento a la evaluación del año anterior: revisar los mecanismos de selección de proyectos y localidades atendidas por el programa (ASM para 2008); revisar la focalización del programa en localidades de alta y muy alta marginación (ASM para 2008); especificar la información que los clubes de migrantes que participan deberán compartir con el programa (ASM para 2009). Sólo la última de éstas recomendaciones fue formalizada y atendida casi en su totalidad.

En la evaluación de 2014 también hallé tres recomendaciones aplicables a las ROP: establecer en las ROP 2016 que se elimina el reembolso de recursos federales en los proyectos productivos, otorgándolos como subsidio completo; establecer en colaboración con el IME en las ROP los criterios y procedimientos de “Toma de Nota” que faciliten la validación en el registro de los clubes de migrantes y sus integrantes; incorporar en las ROP 2016 un párrafo

en el que se mencione que los migrantes pueden solicitar información referente a sus proyectos, así como los argumentos por los que éstos fueron apoyados o rechazados. A pesar de que las tres recomendaciones fueron formalizadas, ninguna fue atendida.

En entrevista, el Dr. Gerardo Esquivel comentó que en su experiencia profesional estos dos programas fueron de los más difíciles de evaluar. Señaló que el 3X1 presentaba problemas importantes en su diseño, mientras que los retos del PDZP eran más bien operativos.

De lo anterior, resalta que el programa con más problemas de diseño fue el que menos ajustes realizó a sus ROP a partir de las evaluaciones. El Dr. Esquivel comentó que al evaluar este programa los funcionarios de Sedesol intentaron negociar el contenido de los informes. El Dr. Esquivel explicó que dicha presión, así como la falta de modificaciones al diseño del 3X1, podría ser consecuencia de que se trata de un programa muy bien valorado y políticamente muy vendible por su intención de vincular y potenciar las remesas que envían emigrantes mexicanos a sus comunidades de origen. Hecho que pudo haber dificultado o desincentivado cambios en la lógica de este programa a pesar de que se han identificado alcances limitados y actividades no precisamente de alto impacto económico y social. Además, la confluencia de recursos puede no ser idónea por reproducir desigualdades (Esquivel, 2016).

En contraparte, el PDZP se trataba de un programa mejor diseñado, con impacto económico y social, pero de menor atractivo mediático y político y que implementaba en mayor medida las observaciones a sus ROP.

(B) EFECTO NO INMEDIATO DE LAS EVALUACIONES EN EL ICADI

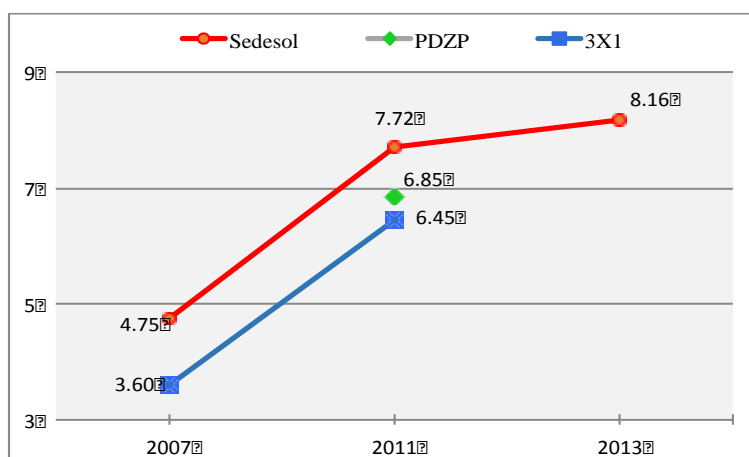
El ICADI estimado por GESOC mostró un avance considerable, entre 2007 y 2011, para todos los programas analizados. Los dos estudios de caso se mantuvieron ligeramente por debajo del promedio obtenido por toda la Secretaría (*véase gráfico 5.1*).

Por un lado, el programa 3X1 mostró un incremento considerable pasando, en cuatro años, de una calificación de 3.6 a 6.45 (en una escala del uno al diez). Mejora que no coincide con el nivel de implementación de las recomendaciones de diseño. Por otro lado, en el caso

del PDZP, fue imposible comparar la evolución de este índice durante el periodo ya que el programa inició sus operaciones en 2009, por lo que sólo estaba disponible una valoración, la de 2011.

Considero que, hasta cierto punto, el ICADI incorpora información relevante para la Matriz de Indicadores para Resultados (MIR) y es por ello que, a pesar de que no observé en el capítulo 4 la implementación de la mayoría de las recomendaciones derivadas de las ED, a mediano plazo los programas muestran una mejoría considerable en este índice. Esto podría ser resultado de un proceso de mejora continua adoptado por los programas, que sería motivado por el hecho de que los datos de la MIR son monitoreada por la Secretaría de Hacienda y Crédito Público (SCHP).

Gráfico 5. 1 El efecto de las recomendaciones de diseño en el ICADI



Fuente: Elaboración propia.

En entrevista, el exfuncionario de Sedesol “B” me confirmó que, efectivamente, los programas presupuestarios de la secretaría experimentaban un proceso paralelo de mejora continua a partir de la introducción del PbR, aproximadamente desde 2009. En este, la SHCP ayudaba, con recomendaciones específicas a los responsables de los programas, a mejorar los indicadores de la MIR, la definición de sus poblaciones, así como la lógica vertical y horizontal del programa.

De lo anterior, concluyo que las mejoras al diseño de los programas, aludiendo específicamente los ajustes en las ROP, no siempre son atribuibles a las ED. No obstante, la valoración del ICADI muestra que a lo largo de los años se han logrado mejoras significativas

en el diseño de la mayoría de los programas evaluados. Esto es atribuible hasta cierto punto a recomendaciones derivadas de EED y otros procesos de mejora que mantienen los programas, impactando positivamente en su diseño. Sin embargo, si bien se reconocen avances, estos no son suficientes debido a que, hoy en día, los programas todavía presentan problemas sustanciales en temas clave, a pesar de que han pasado 10 años de que se institucionalizó la práctica evaluativa. Este hecho se ve reflejado en las EED que, a pesar de que no es su función, señalan continuamente problemas en el diseño de los programas.

5.3.2. LOS CAMBIOS EN EL DESEMPEÑO

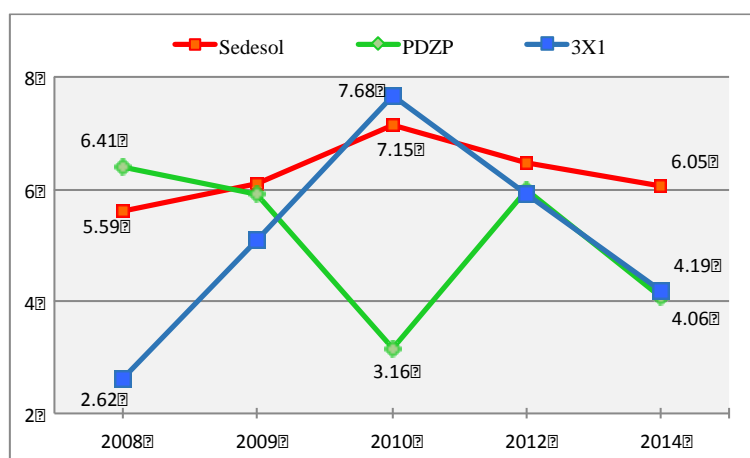
A continuación, analizo el efecto de las evaluaciones en los estudios de caso seleccionados. Para ello tomo en cuenta dos etapas. En la primera, observo la variación del INDEP estimada por GESOC. En la segunda, estudio los cambios en la distribución espacial de los beneficiarios por municipio, empleando como referencia los municipios de alta y muy alta marginación.

Cabe recordar que, en el capítulo 4, el resultado que obtuve al explorar el efecto de las EED en la valoración del INDEP mostró a través de un análisis estadístico que, aunque modesta, sí existe una relación entre las evaluaciones y el desempeño de los programas (*véase ilustración 5.2*). También permitió confirmar que, efectivamente, la utilización de las evaluaciones ha ayudado a los programas de Sedesol a incrementar su capacidad para alcanzar los objetivos establecidos.

(A) EFECTO EN EL INDEP

En términos generales, la evolución registrada en el INDEP de los programas analizados fue heterogénea, lo que me permitió identificar tanto avances como retrocesos (*véase capítulo 2*). En los dos estudios de caso también se presentan tendencias contrarias pero que convergen en 2012 y mantienen la misma tendencia para el 2014 (*véase gráfico 5.2*).

Gráfico 5. 2 El efecto de las evaluaciones en el INDEP



Fuente: Elaboración propia.

El programa 3X1 registró un crecimiento importante entre 2008 y 2010, pasando de una calificación muy baja de apenas 2.62 a 7.68, por encima de la media de Sedesol. Tendencia que se revierte los dos siguientes años, hasta descender a una valoración 4.19 en 2014. Sobre los principales retos operativos de este programa, el exfuncionario de Sedesol “B” señaló que uno era conciliar el tipo de obra que se realizaba con la implementación del mismo, pues la Secretaría promovía las obras de atención prioritaria, pero los emigrantes preferían obras simbólicas para la vida social de la comunidad (ej. La construcción de arcos a la entrada de la ciudad, construcción o remodelación de plazas, fachadas de iglesias o lienzos charros), sobre otras obras de menor relevancia para el desarrollo socio-económico.

Por su parte, el PDZP, inicia con una valoración aceptable de 6.41 y mantiene una caída importante los siguientes años hasta llegar a su mínimo de 3.16, en el año 2010. Posteriormente, para 2012, muestra un progreso que lo posiciona con una valoración próxima a 6 para luego descender nuevamente a 4.06 en 2014. Este programa solo muestra mejoras en la valoración del INDEP un año, en 2012. Esto parece ser reflejo de los problemas operativos que presenta el programa, mismos a los que ya hice alusión antes, a partir de la entrevista con el Dr. Esquivel.

El exfuncionario de Sedesol “B”, señaló que la fusión de programas que dio origen al PDZP implicó un crecimiento presupuestal que podría ser 20 veces mayor al que manejaban con anterioridad los funcionarios. Estos montos superaron la capacidad operativa de los delegados y su conocimiento sobre licitaciones. Con anterioridad, los delegados decidían qué obras realizar y después el gobierno central les señalaba las acciones y dónde se llevarían a cabo. Los municipios también se veían rebasados ya que, debido a los cambios, tenían que gestionar obras cuyo presupuesto podía ser igual o mayor al que ellos recibían anualmente para su localidad. En este sentido, una de las iniciativas implementadas para mejorar sus capacidades fue destinar más recursos a las delegaciones para sus actividades operativas. Sin embargo, el incremento presupuestal realizado sin anticipar las limitaciones operativas fue un gran reto para el programa.

De lo anterior, se puede apreciar que las entrevistas ayudaron a entender mejor los problemas operativos de los dos programas, pero no a explicar los cambios en la valoración del INDEP.

REFLEXIONES FINALES

A manera de conclusión, señalo cinco observaciones derivadas del análisis de los estudios de caso, hallazgos que considero relevantes.

Primero, que la creación de Coneval tuvo dos efectos: por un lado, incrementa la distancia entre evaluador y evaluado, y homogeniza el proceso. Por otro, reduce el protagonismo y la participación de la unidad de evaluación de Sedesol que hasta la fecha había sido clave para la formalización los ASM. Un proceso de formalización que, como ya mencioné, presenta mayor nivel de complejidad al esperado, ya que deriva de negociaciones donde la voluntad del jefe de departamento es determinante y no existen adecuados que lo facilite.

Segundo, en cuanto al desempeño de los evaluadores, cabe señalar que si bien ganan independencia a través de la creación de Coneval, también pierden libertad al reducirse el número de preguntas al que deben responder las evaluaciones y, además, al estar obligados a atender las restricciones que señalan los criterios de los Términos de Referencia.

Tercero, que el seguimiento a la atención de los ajustes o productos derivados de los ASM es un proceso interno de las secretarías. A partir de éste proceso, no queda claro cómo se comprueba la pertinencia o calidad de los ajustes, ya que en la regulación no se establece que los evaluadores deban estar involucrados en esta etapa.

Cuarto, al intentar explicar por qué mejoraron considerablemente los programas en su ICADI, si gran parte de las recomendaciones de diseño que se atienden no impactaron en las ROP, noté algo muy interesante: el proceso de mejora de la MIR, a cargo de la SCHP, parece tener un efecto considerable en el mejoramiento del diseño de los programas y por tanto también en el la valoración del ICADI. Este último podría ser igual o mayor que el de las ED. Para ese aspecto, valdría la pena realizar este análisis en estudios posteriores.

Quinto, por último, en cuanto a la tendencia observada en la valoración del INDEP a lo largo del periodo, las entrevistas solo me permitieron entender los retos que enfrentaban los dos programas analizados, pero no explicar cómo las evaluaciones pudieron vincularse a estos cambios.

Este proceso de formalización, como ya mencioné, presenta mayor nivel de complejidad pues no radica solo en la aceptación (total o parcial) o el rechazo de las recomendaciones emitidas, sino que se involucran otros escenarios: ASM no identificados entre las recomendaciones, repetición de ASM señalados por el mismo evaluador que es contratado de manera consecutiva dos o más años para un mismo programa, repetición de ASM señalados por otro evaluador en informes anteriores y la aceptación aparente de recomendaciones que en la práctica no derivan en acciones de mejora.

CAPÍTULO 6

CONCLUSIONES

CAPÍTULO 6. CONCLUSIONES

Esta tesis respondió al interés de conocer qué tanto han contribuido las evaluaciones realizadas, entre 2007 y 2014, a mejorar los programas de la Secretaría de Desarrollo Social (Sedesol) de México. Además, aporta evidencia sobre su utilización y los beneficios que estas han generado; de manera inmediata en el diseño y, no inmediata, en el desempeño de los programas sociales. Esto con el fin último de explorar si la práctica evaluativa ha logrado su propósito o si requiere ser replanteada.

Presento un análisis exploratorio donde identifico cuál es el alcance o impacto previsto de las evaluaciones en la gestión pública, desde la perspectiva de diferentes actores vinculados al proceso. Retomo la opinión de funcionarios, académicos y evaluadores a partir de entrevistas y declaraciones para entender qué tan útiles se espera que sean las evaluaciones. Esto me permitió precisar que no existe un propósito único y consensado en torno a la práctica evaluativa en México y, por tanto, determinar su nivel de éxito es un ejercicio complejo.

Posteriormente, A partir de los resultados obtenidos en la presente investigación señalo los principales hallazgos que permiten responder a la pregunta, qué tanto contribuyeron las evaluaciones a mejorar el diseño y los resultados de los programas durante el periodo analizado. Adicionalmente, desde un enfoque académico-instrumental señalo los principales retos para la práctica evaluativa orientada a programas sociales.

6.1. SOBRE EL ALCANCE DE LA PRÁCTICA EVALUATIVA

Al investigar qué tanto ayudaron las evaluaciones a mejorar los programas de Sedesol entre 2007 y 2014, asumí, desde un enfoque académico-instrumental (*véase revisión bibliográfica sobre el uso de las evaluaciones*), que el objetivo final de esta práctica, al margen de cualquier otro uso, debería promover mejoras a la gestión pública e indirectamente beneficiar a la sociedad. Sin embargo, debido a que durante mi investigación documental y las entrevistas realizadas noté que este supuesto puede no estar presente en todos los casos, consideré pertinente identificar cuál es la expectativa real que existe entre diversos actores sobre esta práctica. Esta expectativa puede definirse por las atribuciones y las capacidades de las organizaciones que representan, por obvias razones.

Para brindar un panorama general sobre qué esperan de la evaluación los diferentes actores identificados: Secretarías de Estado y organismos descentralizados (SHCP, SFP, Sedesol, Coneval), el Congreso de la Unión y la ASF, así como Organizaciones de la Sociedad Civil (academia y ONG's). En la ilustración 6.1, muestro cómo percibí durante la elaboración de esta tesis la relación esperada entre las evaluaciones y la gestión gubernamental por parte de cada uno de los actores. A pesar de las limitaciones de este apartado, considero que aporta una reflexión que vale la pena compartir.

Para determinar la postura de cada uno de los actores consideré dos dimensiones: por un lado, el tipo de utilización que se promueve a nivel institucional (instrumental, política o conceptual) y, por otro, los beneficios que sus funcionarios esperan a partir de esta práctica (transparencia y rendición de cuentas, mejoras al desempeño de la gestión pública o calidad del gasto) (*véase ilustración 6.1*). Para la segunda dimensión, rescato declaraciones que obtuve a través de entrevistas o ponencias presentadas del *Seminario Internacional sobre la Utilización de las Evaluaciones* que tuvo lugar en mayo de 2016 en la Ciudad de México.

Ilustración 6. 1 Posturas de diferentes actores sobre el tipo de aprovechamiento que debe hacerse de los hallazgos derivados de las evaluaciones

Transparencia y rendición de cuentas	Mejor desempeño de la gestión pública	Calidad del gasto público	
Coneval GESOC	Sedesol SHCP	SHCP	Instrumental
Coneval GESOC Sedesol	SFP ASF	SHCP Cámara de Diputados Comisión de Pto. y Cta. Pública	Político
Coneval GESOC SHCP	Coneval SFP Academia		Conceptual

Fuente: Elaboración propia.

Debido a que el análisis sobre las expectativas de actores es complejo y brinda elementos suficientes para ser objeto de una investigación posterior no entro a más detalle ni intento explicar la complejidad de lo observado, a continuación señalo de manera sucinta cuáles fueron las expectativas identificadas entre los principales actores:

- CONEVAL: La estandarización de la práctica evaluativa, la implementación de las evaluaciones y su difusión;
- SHCP: La implementación y el seguimiento a ASM, con énfasis en la consolidación del PbR;
- SFP: La supervisión del cumplimiento de la normatividad, además estableció un vínculo con la academia para en la promoción de casos exitosos. No obstante, en 2015 deja de tener atribuciones en la materia.
- OSC: El desempeño de los programas y los resultados obtenidos a través de las evaluaciones.
- Académicos y evaluadores: Las restricciones de la actual práctica evaluativa y la profesionalización del campo.

- Sedesol, como dependencia evaluada: La profesionalización de los evaluadores y el seguimiento a las evaluaciones, más como un proceso administrativo que una oportunidad para mejorar su gestión.

6.2. LA UTILIZACIÓN Y EL EFECTO DE LAS EVALUACIONES: BENEFICIOS LIMITADOS EN LOS PROGRAMAS SOCIALES

En primer lugar, cabe mencionar que los resultados de la investigación permitieron validar la hipótesis que planteé al inicio de esta tesis, en consecuencia, puedo afirmar que la práctica evaluativa tuvo un beneficio limitado en el mejoramiento de los programas sociales que estudié. Para respaldar dicha afirmación destaco los siguientes hallazgos en torno a la utilización y el efecto de las evaluaciones.

(A) LA UTILIZACIÓN DE LAS EVALUACIONES

A partir de esta investigación estimo que, en promedio, 58 por ciento de las recomendaciones emitidas por los evaluadores son formalizadas como Aspectos Susceptibles de Mejora (ASM) y que, de ese porcentaje, solo 65 por ciento se implementa (*véase capítulo 2*). Además, en general, se presenta un descenso sostenido en el aprovechamiento de las evaluaciones (tanto en la formalización como en la implementación) que se enfatiza a partir de 2012.

Un elemento, que no anticipaba, resultó ser crucial para entender el descenso sostenido en la utilización de las evaluaciones: el papel que desempeñaba la Unidades de Evaluación de Sedesol, especialmente antes de la creación de Coneval. Esta Unidad era capaz de incidir en la implementación de las recomendaciones “negociando” con los responsables de los programas, en algunos casos con el apoyo de la Unidad de Planeación. Sin embargo, su participación se redujo considerablemente en los últimos años, afectando la práctica evaluativa debido a que Coneval ha carecido de capacidades formales o informales para garantizar o incentivar el aprovechamiento de las evaluaciones que coordina (*véase capítulo 5*).

También encontré que la proporción de recomendaciones utilizadas por Sedesol pueden explicarse parcialmente mediante factores como: la difusión de los informes (resultados recopilados en Google), la intensidad de las evaluaciones (número de evaluaciones acumuladas), el costo de las mismas y la distancia entre evaluador y evaluado (*véase capítulos 3 y 4*).

(B) EL EFECTO DE LAS EVALUACIONES

Las evaluaciones tuvieron efectos limitados, inmediatos y no inmediatos, los cuales pueden apreciarse en los dos tipos de evaluaciones que analicé, diseño y desempeño.

Evaluaciones de Diseño (ED). Este tipo de evaluaciones, con un efecto casi inmediato, ayudaron a los programas a realizar ajustes en sus Reglas de Operación (ROP); sin embargo, estimo que solo 37 por ciento de las recomendaciones de diseño fueron implementadas en las ROP. Por otro lado, llama la atención que una cantidad importante de cambios que se identifican en estos documentos, antes y después de las evaluaciones, no provienen de recomendaciones emitidas por los evaluadores.

Por lo anterior, debido a que los ajustes en ROP atribuibles a las ED no son suficientes para explicar la mejora identificada por GESOC a través de su Índice de la Calidad de Diseño (ICADI) que publicó en 2007 y 2011, decidí analizar en que consistía dicha variación. Descubrí que estos ajustes a las ROP reflejan mejoras en temas estrechamente vinculados con la Matriz de Indicadores para Resultados (MIR) que está a cargo de la SHCP. A diferencia de Coneval, la Secretaría de Hacienda tiene la capacidad para ofrecer incentivos financieros a los evaluados para que realicen ajustes.

En estudios posteriores, valdría la pena explorar qué tan importantes ha sido los ajustes que provienen de procesos de mejora encabezados por la SHCP para explicar la variación del ICADI de programas sociales.

Evaluaciones Específicas de Desempeño (EED). A partir de un análisis estadístico, encontré que la utilización de las EED ayuda a explicar el desempeño de los programas sociales, aunque en una proporción modesta. Considero que sin evidencia suficiente se le confiere demasiada responsabilidad a la evaluación dentro de la gestión pública y, por lo menos en este caso, se puede apreciar que las evaluaciones solo ayudan a explicar parcialmente los cambios que experimentan los programas. Esto no debería sorprendernos si tomamos en cuenta la diversidad de factores y actores que influyen tanto en la toma de decisiones como en la operación de estos programas.

Un fenómeno que observé durante la investigación y para el cual no encontré explicación es la variación del Índice de Desempeño de los Programas Públicos Federales (INDEP) que estima GESOC. El INDEP mostró una mejora significativa los primeros años y después descensos importantes (*véase capítulo 4*); este comportamiento errático podría atribuirse no a un retroceso real en el desempeño de los programas, sino a la modificación constante de los indicadores que se utilizan para valorar el desempeño de los programas. Por tanto, podría asumir que el comportamiento observado es consecuencia de que se eligen nuevos indicadores que reflejan mejor los retos que enfrentan los programas valorados.

EN RESUMEN ¿POR QUÉ LAS EVALUACIONES HAN TENIDO UN BENEFICIO LIMITADO PARA EL MEJORAMIENTO DE LOS PROGRAMAS DE SEDESOL?

Retomando los hallazgos de esta investigación, considero que la práctica evaluativa inició con mucha fuerza en 2007, bajo el supuesto de que las valoraciones serían insumos determinantes para los ajustes presupuestarios que los diputados realizan año con año. Sin embargo, con el paso del tiempo, los actores vinculados a la operación de los programas descubrieron que los informes de evaluación y del desempeño de los programas no condicionan necesariamente su presupuesto, en consecuencia, el compromiso de los evaluados en cuanto a implementar las recomendaciones de los ejercicios de evaluación se ha ido desvaneciendo gradualmente.

A lo anterior cabe agregar que, si bien Sedesol sigue cumpliendo con las evaluaciones que señala el Programa Anual de Evaluación (PAE), estas son menos aprovechadas porque la Unidad de Evaluación que en años anteriores promovía muy activamente la implementación de las recomendaciones emitidas a esta Secretaría, ahora adopta un rol menos relativamente pasivo, cada vez más evidente, que coincide con la creación de Coneval.

Hay que tomar en cuenta que desde 2008, de manera simultánea, la SHCP ha trabajado con dependencias y entidades para construir una MIR que eventualmente permita la implementación de un PbR⁷² que todavía no logra consolidarse en México. El proceso de mejora de la MIR se ha convertido en un tema prioritario para la APF que podría estar

⁷² En los países en desarrollo, sin embargo, el PbR ha enfrentado críticas severas, en particular, por la complejidad de su implementación. Véase, por ejemplo Joyce (2011).

opacando hasta cierto punto, el proceso de retroalimentación de las evaluaciones externas dado a que la SHCP ofrece incentivos que Coneval no puede. Se trata de dos procesos de mejora que apuntan en la misma dirección, no excluyentes, pero que estas dependencias encabezan sin coordinar esfuerzos.

6.3. SOBRE LOS RETOS IDENTIFICADOS EN LA PRÁCTICA EVALUATIVA ORIENTADA A PROGRAMAS SOCIALES

A partir de esta investigación, identifiqué cuatro grandes retos que deben ser atendidos para mejorar la práctica evaluativa en México: 1) las restricciones derivadas de los Términos de Referencia (TR), 2) la limitada capacidad de respuesta de los responsables de los programas valorados, 3) la burocratización de la práctica evaluativa, y 4) la necesidad de reconocer la evaluación como insumo de un proceso de mejora mediante el cual los programas deben alcanzar eventualmente un nivel de madurez que se traduzca en un desempeño efectivo.

1. Las restricciones de los Términos de Referencia.

Las limitaciones que deben superar los TR son de diversa índole: metodológicas, fuentes de información, integración de otros hallazgos y de la opinión profesional de los evaluadores.

- Restricciones metodológicas. Los TR para las ED y las EED analizadas se ciñen a un cuestionario estandarizado donde no es posible incorporar o adicionar criterios que el evaluador pudiera considerar convenientes de acuerdo a su *expertise*. Por tanto, sería conveniente considerar la posibilidad de realizar ajustes que permitan mejorar la valoración de los programas.
- Tipo de información utilizada. Los dos tipos de evaluaciones estudiadas basan su análisis en información oficial que proveen las dependencias valoradas de manera exclusiva. Es importante que las evaluaciones no se limiten a esta fuente ya que de esta manera es imposible valorar la fiabilidad de los datos que reporta el gobierno.
- Integración de otros hallazgos. La estructura de los TR no considera un apartado donde sea posible integrar hallazgos relevantes no previstos en el proceso de valoración. Sería conveniente que los TR brinden a los evaluadores la posibilidad de agregar al análisis este tipo de información.
- Consideración de la opinión crítica de los evaluadores. Otro aspecto que limita la labor de los evaluadores, de acuerdo a las entrevistas, es la imposibilidad de aportar su opinión sobre la valoración realizada y poder realizar recomendaciones. Por lo anterior,

es necesario que los TR cuenten con un apartado donde los evaluadores puedan enriquecer su tarea con un análisis crítico sobre la valoración o el proceso experimentado.

Sin embargo, a pesar de que los evaluadores entrevistados coinciden en la falta de espacios para incorporar cierto tipo de hallazgos en los informes, plasmar su opinión crítica como expertos en la materia y en general, sobre las restricciones que enfrentan al llevar a cabo su trabajo; éstos reconocen que ha existido una buena relación con Coneval ya que cumple su trabajo respaldándolos como intermediario, brindándoles la información requerida o resolviendo dudas.

2. La limitada capacidad de respuesta de los operadores de los programas valorados.

En aquellos casos donde los programas enfrentan retos sustanciales en torno a su diseño o su desempeño, la naturaleza o la complejidad de los Aspectos Susceptibles de Mejora (ASM) pueden superar el poder de decisión de los responsables de los programas. Por lo anterior, es importante involucrar en el proceso a actores con atribuciones aptas para el tipo de retroalimentación requerida y áreas que permitan la consideración de los hallazgos durante la planeación de los programas (esto refiere a que diferentes niveles de la organización se apropien de las evaluaciones).

Usualmente, dentro de las dependencias, la presión por atender las recomendaciones derivadas de las evaluaciones recae en las unidades que operan los programas. Y si estas unidades no cuentan con la capacidad o los recursos para implementar cambios importantes que las evaluaciones recomiendan, pueden optar por comprometer ASM que conllevan ajustes menores (en algunos casos solo aspectos administrativos), más de forma que de fondo. En suma, estas prácticas pueden contribuir a burocratizar la práctica evaluativa.

3. La burocratización de la práctica evaluativa, la falta de incentivos

Durante esta investigación identifiqué dos elementos que sugiero se atiendan para evitar el sub-aprovechamiento o burocratización de las evaluaciones y a su vez, el riesgo de que esta práctica se convierta en un proceso de poco impacto para la gestión pública: la lenta consolidación del PbR que todavía no logra implementarse en México y la falta de otro tipo de incentivos (no presupuestarios), dentro del marco legal vigente, que permitan a Coneval promover el cumplimiento de los ASM.

Ante la falta de incentivos (presupuestarios u otros), las EED se podrían considerar sub-aprovechadas si consideramos que la SHCP y el Congreso deberían utilizarlas para la asignación de recursos de acuerdo con el desempeño de los programas. Sin embargo, hasta ahora, los principales usuarios de estas evaluaciones son los responsables de los programas quienes han empleado los hallazgos para definir ASM.

Aunado a lo anterior, estimo que es necesario analizar la pertinencia de las EED ya que se realizan todos los años pero, como ya mencioné, se nutren exclusivamente de datos oficiales facilitados por las dependencias evaluadas. Por ello, sería conveniente valorar la posibilidad de reorientar los recursos financieros para también realizar evaluaciones que, por ejemplo, mediante trabajo de campo o ejercicios de minería de datos permitan conocer el desempeño de los programas y, después, confrontar con los datos oficiales. Esto significa, en consonancia con la práctica internacional, dejar que en los procesos de seguimiento y monitoreo, se utilicen y difundan los datos gubernamentales.

4. La evaluación como un proceso de mejora que debe alcanzar su objetivo en algún momento y dejar lugar a las tareas de monitoreo

La práctica evaluativa, como proceso de mejora continua, estará presente permanentemente en la gestión pública, sin embargo, su efectividad se aprecia en la medida en que los procesos de evaluación (y retroalimentación) permiten que los programas sociales logren sus objetivos. Esta efectividad debería traducirse, eventualmente, en una práctica evaluativa de menor intensidad que adquiere un papel secundario y ceda protagonismo a las actividades de monitoreo, así como a la implementación de un *Presupuesto basado en Resultados*.

La práctica evaluativa no debe apreciarse solo a través del número de valoraciones o el porcentaje de los ASM atendidos, sino del impacto que tiene para la gestión pública y el logro de los objetivos de los programas. En este sentido, es indispensable que se realicen más estudios sobre el alcance que en términos prácticos han tenido las recomendaciones derivadas de las evaluaciones en el diseño y el desempeño de los programas.

Además de los cuatro retos identificados para la evaluación de programas sociales, hay otros dos aspectos que vale la pena resaltar: la necesidad de fortalecer la cultura de evaluación en general y la percepción sobre el efecto de las evaluaciones en los problemas sociales.

Fortalecimiento de la cultura de evaluación:

En general, observé que se requiere fortalecer la cultura de evaluación en el país para lograr un mayor aprovechamiento de los hallazgos a favor de la gestión pública. Y un aspecto relevante que considero se vincula con el punto anterior, es el debate que gira en torno a la discrecionalidad con que los funcionarios pueden decidir adoptar, o no, las recomendaciones emitidas por los evaluadores. Es decir, se sugiere que los funcionarios (con alusión a los altos mandos) son políticos y como tal deben tomar decisiones políticas que podrían no coincidir con las posiciones de los expertos en evaluación.

Al respecto, considero que es necesario otorgar mayor importancia a esta discusión y centrarse en definir en qué ámbitos es aceptable que criterios políticos rijan decisiones que impactan la gestión pública y qué tipo de recomendaciones deben ser atendidas por considerarse prioritarias para el éxito de los programas, en especial aquellas que podrían poner en riesgo el adecuado desempeño de los programas o el logro de resultados concretos de las intervenciones públicas.

La percepción sobre el efecto de las evaluaciones:

Al margen del éxito con que se ha introducido la práctica evaluativa en México, a partir de su institucionalización, es difícil hablar sobre el efecto que han tenido las evaluaciones en la realidad nacional. Si bien Coneval realiza un esfuerzo importante por aplicar diferentes

evaluaciones a partir de criterios específicos y en un orden determinado (que en un escenario ideal guiarían a los programas a un proceso de mejora que les permitiría cumplir con sus objetivos) sus beneficios para la sociedad no son evidentes. Los problemas sociales que deben atender los programas públicos orientados al combate de la pobreza no muestran una mejora significativa y, a pesar de las evaluaciones a las que son sometidos los programas, persisten áreas de mejora importantes (de diseño o de desempeño). Un logro de la evaluación es que aun cuando no se ha avanzado en estas áreas, por lo menos, ahora son evidentes o se reconocen públicamente.

De lo anterior, surgen las siguientes preguntas: ¿cuántas veces es necesario repetir las evaluaciones para lograr que los programas sean capaces de cumplir sus objetivos explícitos?, y ¿cuál es el alcance real que puede tener esta práctica en la gestión pública?

BIBLIOGRAFÍA

BIBLIOGRAFÍA

- Aarts, K., y Thomassen, J. (2008). Satisfaction with democracy: Do institutions matter? *Electoral Studies*, 27(1), 5–18. doi:10.1016/j.electstud.2007.11.005
- Aguilar Villanueva, L. F. (2007). El aporte de la política pública y la Nueva Gestión Pública a la gobernanza. In *XII Congreso Internacional sobre la Reforma del Estado y de la Administración Pública* (pp. 1–15). CLAD.
- Alkin, M. C., y Christie, C. A. (2004). An Evaluation Theory Tree. In M. C. Alkin (Ed.), *Evaluation Roots: Tracing Theorists' Views and Influences* (pp. 12–65). Sage Publications, Inc.
- Anderson, C. J. (1998). Political Satisfaction in Old and New Democracies. *Department of Political Science*. Binghamton University.
- Balthasar, A. (2009). Institutional Design and Utilization of Evaluation on Swiss Experience, 226–256.
- Banco Mundial. (2009). *El Sistema de MyE de México: Un salto del nivel sectorial al nacional*.
- Barnette, J. J., & Thompson, J. C. (1979). A descriptive assessment of the effect of evaluations on instruction. *Studies in Educational Evaluation*, 5(1), 77-86.
- Calvo, E. G. (2010). Calidad democrática: enfoque teórico y posible medición. In *Calidad Democrática y Buen Gobierno* (pp. 28–52). FEMP - AEVAL.
- Cardozo Brum, M. (2009). La institucionalización de una cultura de la evaluación en la administración pública mexicana : avances y desafíos pendientes.
- Cejudo, G. M., y Maldonado Trujillo, C. (Eds.). (2011). *De las recomendaciones a las acciones: La experiencia del Premio 2011. Programas Federales Comprometidos con el Proceso de Evaluación*. CIDE, SFP y CLEAR.

- Centro Latinoamericano de Administración para el Desarrollo (CLAD) (2007). El Sistema Nacional de Monitoreo y Evaluación de la Gestión Pública en México. Banco Mundial. En: <http://siare.clad.org/siare/innotend/evaluacion/mexico.pdf>
- Consejo Nacional de Evaluación de la Política de Desarrollo Social (Coneval) (2016). La Política de evaluación en México: 10 años del CONEVAL. México.
- Cousins, J. B. (2004). Commentary: Minimizing Evaluation Misuse as Principled Practice, 25(3), 391–397.
- Cousins, J. B., y Leithwood, K. A. (1986). Current Empirical Research on Evaluation Utilization (1971-1985). *Review of Educational Research*, 56(3), 331–364. doi:10.3102/00346543056003331
- Delgado Sotillos, I. (2015). Instituciones y satisfacción con la democracia. Un estudio comparado de los vínculos representativos en Latinoamérica. *Revista latinoamericana de política comparada*. Centro Latinoamericano de Estudios Políticos. Retrieved from <http://dialnet.unirioja.es/servlet/articulo?codigo=4959324>
- Delgado Sotillos, Irene (2015). Instituciones y satisfacción con la democracia. Un estudio comparado de los vínculos representativos en Latinoamérica. *Revista Latinoamericana de Política Comparada, CELAEP*.
- Diario Oficial de la Federación (dof) (2005). Decreto por el que se regula el consejo nacional de evaluación de la política de desarrollo social. Gobierno de México.
- Fernando Castro, M., Lopez-Acevedo, G., Beker Busjeet, G., y Fernandez Ordonez, X. (2009). El Sistema de MyE de México: Un salto del nivel sectorial al nacional. *Serie de Documentos de Trabajo DCE N° 20 - Desarrollo de La Capacidad de Evaluación*. Banco Mundial - Independent Evaluation Group. Retrieved from http://www.coneval.gob.mx/rw/resource/coneval/info_public/El_sistema_de_monitoreo_de_Mexico_esp.pdf
- García López, R., y García Moreno, M. (2010). *La Gestión para Resultados en el Desarrollo. América Latina y el Caribe* (2ª ed.). BID.

- Gray, Cherly (2016). Presentación: “Conocimiento sobre el uso de la evaluación”. Seminario Internacional de estudios empíricos sobre la utilización de la evaluación. 24 de mayo 2016. Disponible en: https://www.youtube.com/watch?v=Cso5_Iec2Kc
- Gris, C. (2014). El uso de la evaluación como herramienta para la innovación gubernamental: el caso del programa oportunidades. *XXVII Concurso del CLAD sobre Reforma del Estado y Modernización de la Administración Pública "La Innovación en la Gestión Pública"*.
- Henry, G. T., y Mark, M. M. (2003). Toward an Agenda for Research on Evaluation, (97), 69–80.
- Herbert, J. L. (2014). Researching evaluation influence: a review of the literature. *Evaluation Review*, 38(5), 388–419. doi:10.1177/0193841X14547230
- House, E. (1992). Tendencias en evaluación. *Revista de Educación*, (299), 43–56.
- Hofstetter, C. H., and Alkin, M. (2003). Evaluation use revisited. In Kellaghan T, Stufflebeam DL, Wingate L. (Eds.), *International handbook of education evaluation*.
- Husseiny, N. y Earl, S. (1996). Enhancing the use of Evaluation Findings in Organizations
- Joyce, P. G. (2011), The Obama Administration and PBB: Building on the Legacy of Federal Performance-Informed Budgeting?. *Public Administration Review*, 71: 356–367. doi:10.1111/j.1540-6210.2011.02355.x
- Johnson, K., Greenesid, L. O., Toal, S. A., King, J. A., Lawrenz, F., y Volkov, B. (2009). Research on Evaluation Use: A Review of the Empirical Literature From 1986 to 2005. *American Journal of Evaluation*, 30(3), 377–410. doi:10.1177/1098214009341660
- Kaufmann, D., Kraay, A., y Mastruzzi, M. (2007). *Governance Matters VI: Governance Indicators for 1996-2006. Policy Research Working Paper No. 4280*. Retrieved from <http://ssrn.com/abstract=999979>

- Kaufmann, D., y Kraay, A. (2008). *Governance Indicators: Where Are We, Where Should We Be Going? Policy Research Working Paper No. 4370*. University Library of Munich, Germany. Retrieved from <http://ideas.repec.org/p/pramprapa/8212.html>
- Kirkhart, K. E. (2000). *Reconceptualizing Evaluation Use: An Integrated Theory of Influence*, (88).
- Ledermann, S. (2012). Exploring the Necessary Conditions for Evaluation Use in Program Change. *American Journal of Evaluation*, 33(2), 159–178. doi:10.1177/1098214011411573
- Leviton, L. y Boruch, R.F. (1983). *Contributions of evaluation to education programs and policy. Evaluation Review* 7/5.
- Montero, J., Zmerli, S., y Newton, K. (2008). Confianza social, confianza política y satisfacción con la democracia. *Revista Española de Investigaciones Sociológicas (REIS)*, 122, 11–54.
- Morales Quiroga, M. (2009). Corrupción y democracia. América Latina en perspectiva comparada. *Gestión y Política Pública*, XVIII(2), 205–252. Retrieved from <http://www.redalyc.org/articulo.oa?id=13311978001>
- Morlino, L. (2004). What is a “good” democracy? *Democratization*, 11(5), 10–32. doi:10.1080/13510340412331304589
- Ospina Bozzi, S. (2002). Construyendo capacidad institucional en América Latina: el papel de la evaluación como herramienta modernizadora. In *VII Congreso Internacional sobre la Reforma del Estado y de la Administración Pública* (p. 21). Lisboa: CLAD.
- Patiño, Mireya (2014). La utilización efectiva de las evaluaciones en las decisiones sobre política pública. El caso de los programas federales de atención a grupos vulnerables. Tesis para obtener el grado de Maestría. El Colegio de México.
- PNUD. (2008). *Desarrollo de Capacidades. Nota de Práctica*. Retrieved from http://www.undp.org/content/dam/undp/library/capacity-development/spanish/SP_Nota de Practica_Desarrollo de Capacidades.pdf

- PNUD. (2009). Manual de Planificación, Seguimiento y Evaluación de los Resultados de Desarrollo. Retrieved from <http://www.undp.org/eo/handbook>
- Reichardt, Ch. y Cook, T (2011). Métodos cualitativos y cuantitativos de investigación.
- Rojas Aravena, F. (2004). Democracia y gobernabilidad en América Latina. *Papeles de Cuestiones Internacionales*, Verano (86, 109–120.
- Befani B., Ledermann S., Sager F. (2007). ... (2008). Advances in mixed methods research. London, Thousand Oaks, CA, New Delhi, and Singapore: Sage.
- Shulha, L. y Cousins, B. (1997). Evaluation use: theory, research and Practice Since 1986
- Taut, S. M., y Alkin, M. C. (2003). American Journal of Evaluation. doi:10.1177/109821400302400205
- Torres-Rivas, E. (1993). América Latina. Gobernabilidad y democracia en sociedades en crisis. *Nueva Sociedad* Nro. 128 Noviembre-Diciembre, pp. 88-101.
- Vargas Chanes, D., y González Núñez, J. C. (2013). Los determinantes de la satisfacción con la democracia, desde el enfoque de un modelo multinivel. *EconoQuantum*, 10(2), 55–75. Retrieved from <http://www.redalyc.org/articulo.oa?id=125029955003>
- Wagner, A. F., Schneider, F., y Halla, M. (2009). The quality of institutions and satisfaction with democracy in Western Europe - A panel analysis. *European Journal of Political Economy*, 25(1), 30–41. doi:10.1016/j.ejpoleco.2008.08.001
- Weiss, C. (1998). Have we learned anything new about the use of evaluation? *The American Journal of Evaluation*, 19(1), 21–33. doi:10.1016/S1098-2140(99)80178-7
- Weiss, C. H. (1972). *Evaluation research: methods for assessing program effectiveness*. Prentice-Hall.
- Wollmann, H. (2003). *Evaluation in public-sector reform: Toward a “third wave” of evaluation?*

Bases de datos utilizadas:

Banco Mundial [en línea]: [fecha de consulta, mayo 2015]. Base de datos disponible sobre los Indicadores mundiales de Gobernanza 1996-2013. Disponible en:

<<http://info.worldbank.org/governance/wgi/index.asp>>

CONEAL [en línea]. Ciudad de México: [fecha de consulta, marzo 2015]. Base de datos sobre la Medición de la Pobreza en México 2010-2014. Disponible en:

<http://www.coneval.gob.mx/Medicion/MP/Paginas/Pobreza_2014.aspx>

GESOC [en línea]. Ciudad de México: [fecha de consulta, marzo 2015]. Base de datos sobre el Índice de Desempeño de Programas Públicos Federales 2014. Disponible en:

<<http://www.indep.gesoc.org.mx/>>

Latinobarómetro [en línea]. Santiago, Chile: [fecha de consulta, mayo 2015]. Base de datos sobre el Grado de satisfacción con el funcionamiento de la democracia 1995-2013. Disponible en: <<http://www.latinobarometro.org/latOnline.jsp>>

Evaluaciones realizadas a programas de Sedesol entre 2006 y 2012. Disponibles en:

[http://www.2006-2012.sedesol.gob.mx/es/SEDESOL/Evaluacion de Los Programas Sociales](http://www.2006-2012.sedesol.gob.mx/es/SEDESOL/Evaluacion_de_Los_Programas_Sociales)

Avances Susceptibles de Mejora Derivados de Evaluaciones e Informes de programas de Sedesol para el periodo 2006-2012. Disponible en:

<[http://www.2006-2012.sedesol.gob.mx/es/SEDESOL/Aspectos Susceptibles de Mejora Derivados de Evaluaciones e Informes](http://www.2006-2012.sedesol.gob.mx/es/SEDESOL/Aspectos_Susceptibles_de_Mejora_Derivados_de_Evaluaciones_e_Informes)>

Reglas de Operación de los Programas de Sedesol para el periodo 2006-2012. Disponible en:

<<http://padrones.inmujeres.gob.mx/listaProgramas.php>>

ANEXOS

ANEXOS

Preguntas para las evaluaciones de diseño 2007-2008

Contribución a los objetivos estratégicos de Sedesol y nacionales
1. Problema o necesidad prioritaria
2. Existencia de un diagnóstico actualizado y adecuado sobre la problemática
3. Fin y Propósito claramente definidos
4. Fin y Propósito a la solución del problema
5. Evidencia de estudios o investigaciones nacionales
6. Vinculo con objetivos estratégicos de sedesol
7. Relación lógica con los objetivos del PND
Lógica vertical de la MIR
8. Actividades suficientes y necesarias para producir Componentes.
9. Componentes necesarios y suficientes para el logro del Propósito
10. Contribución del propósito al logro del Fin
11. Validez de la lógica vertical
12. Cambios en el diseño lógica vertical
Lógica horizontal de la MIR
13. Indicadores para medir el desempeño del programa
14. Características de los indicadores
15. Recomendaciones del evaluador sobre indicadores
16. Línea base y temporalidad de los indicadores
17. Identificación de los medios de verificación
18. Medios de verificación y tamaño de la muestra
19. Veracidad de la información

20. Validez de los supuestos
21. Validez de la lógica horizontal
22. Cambios en el diseño lógica horizontal
Población potencial y objetivo
23. Población potencial/objetivo
24. Cuantificación y caracterización de poblaciones
25. Justificación de los beneficios
26. Conveniencia de la justificación
27. Idoneidad de los criterios y mecanismos utilizados
28. Información sistematizada y actualizada que permita conocer quienes reciben apoyos
Vinculación de las ROP con los objetivos del programa
29. El diseño en las ROP
30. Congruencia entre ROP y lógica interna
Coincidencias, complementariedades o duplicidades de acciones con otros programas federales
31. El diseño para alcanzar el propósito y para atender la población objetivo
32. Complementariedad y/o sinergia con programas federales
33. Duplicidad con programas federales
34. Información sobre complementariedad o duplicidad

Preguntas para las evaluaciones de diseño 2013

Justificación de la creación y del diseño del programa

1. El problema o necesidad prioritaria que busca resolver el programa está identificado en un documento que cuenta con la siguiente información:

- a) El problema o necesidad se formula como un hecho negativo o como una situación que pueda ser revertida.
- b) Se define la población que tiene el problema o necesidad.

2. Existe un diagnóstico del problema que atiende el programa que describa de manera específica:

- a) Causas, efectos y características del problema.
- b) Cuantificación y características de la población que presenta el problema.
- c) Ubicación territorial de la población que presenta el problema.

3. ¿Existe justificación teórica o empírica documentada que sustente el tipo de intervención que el programa lleva a cabo?

Contribución del programa a los objetivos nacionales y los sectoriales

4. El propósito del programa está vinculado con los objetivos del programa sectorial, especial o institucional considerando que:

- a) Existen conceptos comunes entre el propósito y los objetivos del programa sectorial, especial o institucional por ejemplo: población objetivo
- b) El logro del propósito aporta al cumplimiento de alguna(s) de la(s) meta(s) de alguno(s) de los objetivos del programa sectorial, especial o institucional.

5. ¿Con cuáles ejes temáticos y objetivos del Plan Nacional de Desarrollo vigente está vinculado el objetivo sectorial relacionado con el programa?
6. ¿Cómo está vinculado el propósito del programa con las Metas del Milenio?
Población potencial y objetivo
7. Las poblaciones, potencial y objetivo, están definidas en documentos oficiales y/o en el diagnóstico del problema y cuentan con la siguiente información y características: Unidad de medida Están cuantificadas Metodología para su cuantificación Fuentes de información
<p>8. Existe información que permita conocer quiénes reciben los apoyos del programa (padrón de beneficiarios) que:</p> <p>a) Incluya las características de los beneficiarios establecidas en su documento normativo.</p> <p>b) Incluya el tipo de apoyo otorgado</p> <p>c) Esté sistematizada</p> <p>d) Cuente con mecanismos documentados para su depuración y actualización</p>
9. Si el programa recolecta información socioeconómica de sus beneficiarios, explique el procedimiento para llevarlo a cabo, las variables que mide y la temporalidad de las mediciones

Evaluación y análisis de la Matriz de Indicadores para Resultados

10. Para cada uno de los componentes de la MIR del programa existe una o un grupo de actividades que:

- a) Están claramente especificadas, es decir, no existe ambigüedad en su redacción.
- b) Están ordenadas de manera cronológica
- c) Son necesarias, es decir, ninguna de las actividades es prescindible para producir los componentes.
- d) Su realización genera junto con los supuestos en ese nivel de objetivos los componentes.

11. Los componentes señalados en la MIR cumplen con las siguientes características:

- a) Son los bienes o servicios que produce el programa
- b) Están redactados como resultados logrados, por ejemplo becas entregadas
- c) Son necesarios, es decir, ninguno de los componentes es prescindible para producir el propósito
- d) Su realización genera junto con los supuestos en ese nivel de objetivos el propósito

12. El propósito de la MIR cuenta con las siguientes características:

- a) Es consecuencia directa que se espera ocurrirá como resultado de los componentes y los supuestos a ese nivel de objetivos
- b) Su logro no está controlado por los responsables del programa
- c) Es único, es decir, incluye un solo objetivo
- d) Está redactado como una situación alcanzada, por ejemplo: morbilidad en la localidad reducida

e) Incluye la población objetivo

13. El fin de la MIR cuenta con las siguientes características:

- a) Está claramente especificado, es decir, no existe ambigüedad en su redacción
- b) Es un objetivo superior al que el programa contribuye, es decir, no se espera que la ejecución del programa sea suficiente para alcanzar el fin
- c) Su logro no está controlado por los responsables del programa
- d) Es único, es decir, incluye un solo objetivo
- e) Está vinculado con objetivos estratégicos de la dependencia o del programa sectorial.

14. ¿En el documento normativo del programa es posible identificar el resumen narrativo de la MIR (fin, propósito, componentes y actividades)?

15. En cada uno de los niveles de objetivos de la MIR del programa (fin, propósito, componentes y actividades) existen indicadores para medir el desempeño del programa con las siguientes características:

- a) Claros
- b) Relevantes
- c) Económicos
- d) Monitoreables
- e) Adecuados

16. Las Fichas Técnicas de los indicadores del programa cuentan con la siguiente información:

- a) Nombre
- b) Definición
- c) Método de cálculo
- d) Unidad de medida
- e) Frecuencia de medición
- f) Línea de base
- g) Metas
- h) Comportamiento del indicador (ascendente, descendente, regular o nominal)

17. Las metas de los indicadores de la MIR del programa tienen las siguientes características:

- a) Cuentan con unidad de medida.
- b) Están orientadas a impulsar el desempeño, es decir, no son laxas.
- c) Son factibles de alcanzar considerando los plazos y los recursos humanos y financieros con los que cuenta el programa.

18. Cuántos de los indicadores incluidos en la MIR tienen especificados medios de verificación con las siguientes características:

- a) Oficiales o institucionales
- b) Con un nombre que permita identificarlos
- c) Permiten reproducir el cálculo del indicador

d) Públicos, accesibles a cualquier persona

19. Considerando el conjunto Objetivo-Indicadores-Medios de verificación, es decir, cada renglón de la MIR del programa es posible identificar lo siguiente:

a) Los medios de verificación son los necesarios para calcular los indicadores, es decir, ninguno es prescindible.

b) Los medios de verificación son suficientes para calcular los indicadores.

c) Los indicadores permiten medir, directa o indirectamente, el objetivo a ese nivel.

20. Modificaciones a la MIR del programa o incorpore los cambios que resuelvan las deficiencias encontradas en cada uno de los elementos a partir de sus respuestas a las preguntas del apartado

Posibles complementariedades y coincidencias con otros programas federales

21. ¿Con cuáles programas federales y en qué aspectos el programa evaluado podría tener complementariedad y/o coincidencias?

SELECCIÓN DE CASOS: LAS EVALUACIONES ANALIZADAS

Los Lineamientos Generales para la Evaluación de los Programas Federales de la Administración Pública Federal (de aquí en adelante Lineamientos Generales) definen siete tipos de evaluaciones aplicables;⁷³ de éstas, sólo analizo dos que cuentan con los índices emitidos por GESOC para estimar la capacidad de los programas evaluados para alcanzar sus objetivos, estas son: las de *diseño* y las de *desempeño* (véase apartado 1.2).

Cabe mencionar que los dos tipos de evaluaciones especificados son subcategorías de las siete que se mencionan originalmente los Lineamientos Generales. Por un lado, las de *diseño*, derivan del primer apartado de las evaluaciones de Consistencia y Resultados; por otro, las de *desempeño*, son una variedad de las evaluaciones Específicas.

Para el periodo 2007-2014, CONEVAL reporta que los programas federales de Sedesol estuvieron sujetos a un total de 134 evaluaciones de diseño y desempeño.⁷⁴ De éstas, empleo 111 que cuentan con la valoración de GESOC: 36 de diseño y 75 de desempeño (véase *cuadro 1*).

⁷³ Para garantizar la evaluación orientada a resultados y retroalimentar el Sistema de Evaluación del Desempeño, los Lineamientos Generales, publicados en el DOF el 30 de marzo de 2007, definen siete tipos de evaluaciones aplicables a los programas federales: Consistencia y Resultados, Indicadores, Procesos, Impacto, Específica, Estratégica, Complementarias (mencionados en los numerales décimo sexto, décimo octavo, y vigésimo cuarto).

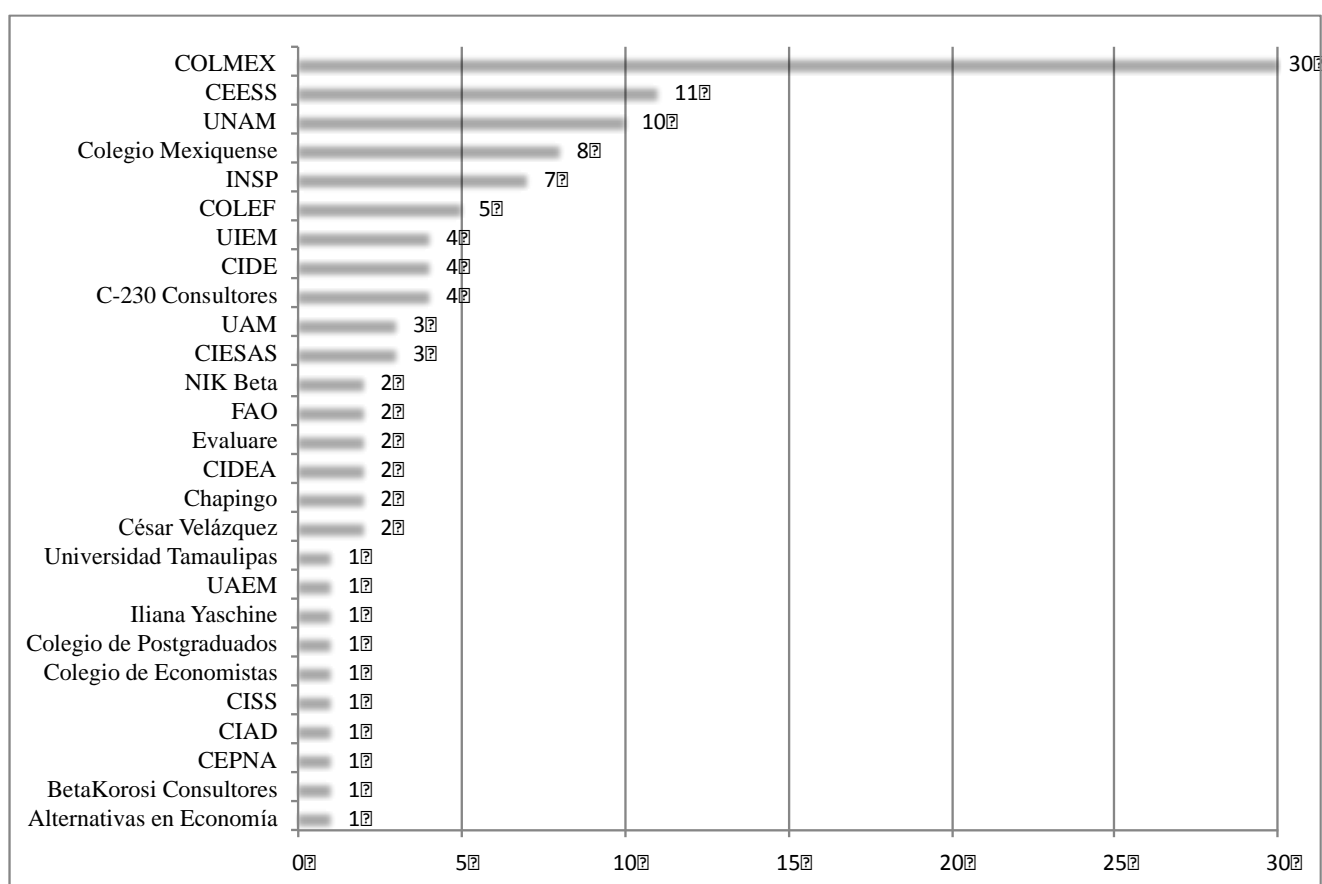
⁷⁴ En cumplimiento de los Lineamientos Generales y al Decreto de Presupuesto de Egresos de la Federación, se publica el Programa Anual de Evaluación para dar a conocer las evaluaciones que se llevarán a cabo o comenzarán durante cada uno de los ejercicios fiscales y a los programas federales que aplican (CONEVAL, 2005 en: <http://www.coneval.gob.mx/Evaluacion/NME/Paginas/ProgramaAnual.aspx>).

Cuadro 1 Reportes de evaluación difundidos por Coneval para el periodo 2007-2014

No.			Nombre del Programa	EED		ED		Total
				CInd	SInd	CInd	SInd	
1	B	004	Programa de Adquisición de Leche Nacional				1	1
2	S	005	Programa de Apoyo Alimentario en Zonas de Atención Prioritaria				1	1
3	U	008	Subsidio a Programas para Jóvenes	1				1
4	S	048	Programa Hábitat	3		2		5
5	S	052	Programa de Abasto Social de Leche a cargo de Liconsa, S.A. de C.V.	5		2		7
6	S	053	Programa de Abasto Rural a cargo de Diconsa, S.A. de C.V.	4	1	2		7
7	S	054	Programa de Opciones Productivas	4	1	2		7
8	S	055	Programa Jóvenes por México				1	1
9	S	057	Programas del Fondo Nacional de Fomento a las Artesanías	5		2		7
10	S	058	Programa de Ahorro, Subsidio y Crédito para la Vivienda "Tu Casa"	3		2		5
11	S	059	Programa para el Desarrollo Local (Microrregiones)	1			1	2
12	S	061	Programa 3x1 para Migrantes	5		2		7
13	S	065	Programa de Atención a Jornaleros Agrícolas	5		2		7
14	S	070	Programa de Coinversión Social	4	1	2		7
15	S	071	Programa de Empleo Temporal	3		1	1	5
16	S	072	Programa de Desarrollo Humano Oportunidades / PROSPERA	4	1	1	1	7
17	S	117	Programa de Vivienda Rural	2	1	2		5
18	S	118	Programa de Apoyo Alimentario	5		2	1	8
19	S	155	Programa de Apoyo a las Instancias de Mujeres en las Entidades Federativas para Implementar y Ejecutar Programas de Prevención de la Violencia Contra las Mujeres	4	1	1	1	7
20	S	174	Programa de Estancias Infantiles para Apoyar a Madres Trabajadoras	4	1	2		7
21	S	175	Programa de Rescate de Espacios Públicos	1	2	2	1	6
22	n.d.	n.d.	Programa Atención a Adultos Mayores en Zonas Rurales			1		1
23	S	176	Programa 70 y Más / Programa Pensión para Adultos Mayores	5		3		8
24	S	213	Programa de Apoyo a los Avecindados en Condiciones de Pobreza Patrimonial para Regularizar Asentamientos Humanos Irregulares	2	1	1	1	5
25	S	216	Programa para el Desarrollo de Zonas Prioritarias	4		1		5
26	S	237	Programa de Prevención de Riesgos en los Asentamientos Humanos				1	1
27	S	241	Programa Seguro de Vida para Jefas de Familia	1		1		2
28	S	716	Programa de Apoyo a Zonas de Atención Prioritaria		1		1	2
				75	11	36	12	134

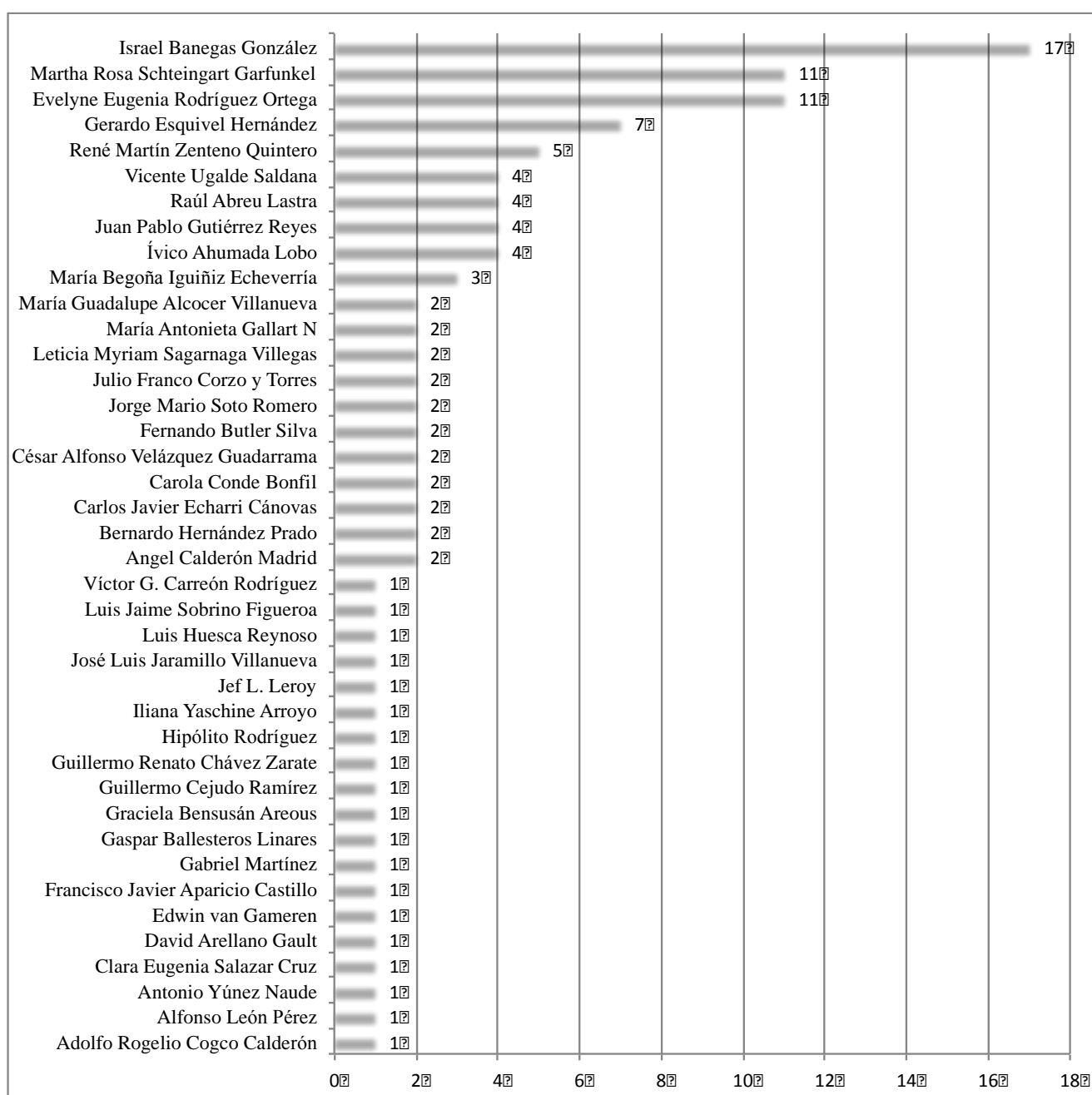
Fuente: Elaboración propia.

Ilustración 1 Número de evaluaciones realizadas por organismo evaluador durante el periodo 2007-2014



Fuente: Elaboración propia.

Ilustración 2 Número de evaluaciones realizadas por coordinador durante el periodo 2007-2014



Fuente: Elaboración propia.

ANÁLISIS DE LA VARIABLE DEPENDIENTE: LA VALORACIÓN DE GESOC

En este apartado defino y explico brevemente la metodología a través de la cual se construyen los índices que utilizo como variable dependiente en esta investigación.

Para los Programas presupuestarios (sujetos a reglas de operación) que son evaluados a partir del 2007, GESOC estima dos índices que permiten valorar su progreso: el Índice de Calidad de Diseño (ICADI), y el Índice de Desempeño de los Programas Públicos Federales (INDEP). Estos índices permiten aprovechar la información generada por las evaluación externa a programas federales, así como fortalecer la rendición de cuentas y el debate público sobre la política gubernamentales. Son una herramienta de fácil acceso que ordena, sistematiza, compara y facilita la difusión de los avances de los programas (GESOC, 2007&2008).

Índice de Calidad del Diseño (ICADI)

El ICADI se basa en la premisa de que “un programa público que cuenta con un diseño claro y coherente, con una visión estratégica de mediano plazo, una operación consistente y un perfil de orientación a resultados y a la ciudadanía, tendrá mayores probabilidades de tener un desempeño satisfactorio y ofrecer mejores resultados vinculados a los objetivos inicialmente propuestos” (ICADI, 2008).

Para su construcción en 2007, los resultados de la ECR fueron codificados y agregados en tres Subíndices que resumen los factores clave de la calidad de diseño de un programa público: Diseño y Alineación Estratégica, de Operación, y de Orientación a Resultados y a la Ciudadanía (véase cuadro2).

Cuadro 2 Ajustes realizados al método de cálculo del ICADI entre 2007 y 2013

Año	Fórmula	Subíndices / Componentes
2007	SubÍndice de diseño y alineación estratégica + SubÍndice de operación + SubÍndice de orientación a resultados y a la ciudadanía	SubDyAE. Definición de objetivos, metas y alcances estratégicos del programa federal evaluado.
		SubO. Previsiones institucionales y organizaciones que permitirían una correcta operación de los programas por parte de las dependencias ejecutoras.
		SubORyC. Previsiones de los programas para medir resultados e impactos, así como la percepción y opinión de la población objetivo.
2011 & 2013	SubÍndice de justificación del programa y contribución a objetivos + SubÍndice de población potencial y objetivo + SubÍndice de Matriz de Indicadores para Resultados (MIR)	SubJ&C. Justificación de la creación y el diseño del programa, así como de la contribución a las metas sectoriales y PND.
		SubPob. Definición e identificación de la población bajo los supuestos del problema público a atender.
		SubMIR. Previsiones relativas a diferentes niveles de objetivos del programa, su calidad y suficiencia.

Fuente: Nota metodológica de ICADI.

El ICADI 2007 estima la calidad de diseño de cada programa en una escala de cero a diez, donde cero es el mínimo y diez representa la excelencia (ICADI, 2008). Además, aplica un sistema de semaforización a los resultados de cada Subíndice para entender mejor la situación

de cada programa, estableciendo los siguientes criterios: las notas entre 10 y 8 son “Destacado”, las que oscilan entre 7.9 y 6 “Suficiente”, y las inferiores a 6 “Rezagado”.

Índice de Calidad de Diseño (ICADI) 2007

Se sujetó a una metodología uniforme estandarizada en un cuestionario de 100 preguntas que debía ser respondido por los evaluadores externos. Del total de preguntas, 80 estaban bajo un formato binario (esto es, aceptaban una respuesta sí/no) y el resto tenían un esquema de respuesta libre. El cuestionario agrupaba las preguntas en seis secciones o capítulos que atendían a cuestiones diversas de la consistencia de los programas federales:

Diseño:

- ☐ Grado de alineación de sus objetivos con los objetivos estratégicos.
- ☐ Adecuación de la matriz de indicadores.
- ☐ Duplicidades del programa.
- ☐ Adecuación en el enfoque, definición del problema y de los objetivos del programa.

Planeación Estratégica:

- ☐ Existencia y adecuación de los mecanismos para orientar la operación de corto, mediano y largo plazo del programa.
- ☐ Existencia y adecuación de los instrumentos de monitoreo y evaluación que auxilian a los implementadores en la orientación a resultados.

Operación:

- ☐ Adecuación del marco legal que sustenta y da vigencia a los programas.
- ☐ Adecuaciones institucionales necesarias para mejorar la operación de los programas.
- ☐ Adecuación de prácticas financieras en la ejecución de programas.
- ☐ Existencia de mecanismos de aprendizaje organizacional.

- ☐ Eficiencia, eficacia y economía operativa del programa.
- ☐ Existencia de mecanismos para la sistematización y actualización de información.

Cobertura y Focalización:

- ☒ Adecuación en la definición y atención de las poblaciones potencial y objetivo del programa.

Percepción de los Beneficiarios:

- ☒ Existencia de mecanismos al interior del programa que permitan medir el grado de satisfacción de los beneficiarios del programa.

Resultados:

- ☒ Existencia de instrumentos que permitan medir el nivel de cumplimiento de los objetivos del programa a nivel “Propósito y Fin” de las matrices de indicadores.

Una vez construida la base de datos se optó por construir tres subíndices básicos mediante las cuales se podían clasificar aspectos diferenciados de la calidad del diseño de un programa:

Subíndice de diseño y alineación estratégica: preguntas relacionadas con la definición de objetivos, metas y alcances estratégicos del programa federal evaluado (**46%** de las preguntas, capítulos de diseño y planeación estratégica).

Subíndice de operación: preguntas relacionadas con las previsiones institucionales y organizacionales que permitirían una correcta operación de los programas por parte de las dependencias ejecutoras (**40%** de las preguntas de los reportes de consistencia y resultados y correspondientes exclusivamente al capítulo de operación).

Subíndice de orientación a resultados y a la ciudadanía: preguntas referentes a las previsiones de los programas para medir resultados e impactos, así como la percepción y opinión de la población objetivo (**14%** de las preguntas de las evaluaciones de consistencia y resultados, capítulos de especial relevancia: cobertura y focalización, percepción de los beneficiarios y resultados).

$$\text{ICADI} = (\text{SubDyAE} + \text{SubO} + \text{SubORyC})$$

SubDyAE= $\sum_{i=0} ((\text{Diseño} + \text{Planeación estratégica} / 37) * 0.45) * 10$, el rango esperado es de 0 a 4.5

SubO= $\sum_{i=0} ((\text{Operación}/32) * 0.30) * 10$, el rango esperado es de 0 a 3

SubORyC= $\sum_{i=0} ((\text{Consistencia y Resultados} + \text{Percepción de beneficiarios} + \text{Relevancia y focalización} / 37) * 0.25) * 10$, el rango esperado es de 0 a 2.5

Semaforización

De 8 a 10, Destacado

De 6 a 7.9, Suficiente

Inferior a 6, Rezagado

Índice de Desempeño de los Programas Públicos Federales (INDEP)

La construcción del INDEP parte de una premisa simple: “todo programa gubernamental que se precie de contar con un buen desempeño debe contar con al menos los siguientes atributos:

1. Un marco de evaluación comprehensivo y completo que permita analizar los avances de un programa en toda la cadena lógica del desempeño (insumos, procesos, productos, resultados e impactos)
2. Metas claras y exigentes que permitan medir el esfuerzo potencial de una institución pública
3. Los avances de sus indicadores deben mostrar avances significativos en la consecución de los objetivos de los programas” (GESOC, 2008).

A partir de la premisa mencionada, en el periodo 2008-2009 el INDEP se estimó con base en dos componentes fundamentales: el Componente de Avance en los Indicadores reportados (CAI) y; el Componente de Fortaleza del sistema de Indicadores (CFI). El primero se centra en el análisis de los avances en las metas observadas, pero castiga comportamientos sospechosos que pudieran revelar un diseño *ad hoc* o estratégico de metas. El segundo evalúa la fortaleza del sistema de indicadores (su alcance) y el marco de evaluación mediante el cual las metas de los programas son medidas. Dado que ambos componentes se consideran atributos necesarios de un programa que muestra buen desempeño, GESOC optó por concebir un índice multiplicativo que obliga a contar con resultados satisfactorios en ambos componentes para otorgar una alta calificación (véase *anexo metodológico*).

Cuadro 1. 4 Ajustes realizados al método de cálculo del INDEP entre 2007 y 2014

Año	Fórmula	Componentes
2008 - 2009	$CAI \times CFI$	CAI= El Componente de Avance en los Indicadores reportados CFI= EL Componente de Fortaleza del sistema de Indicadores
2010	$(CAI + CCPB^2) / 2$	CAI= El Componente de Avance en los Indicadores reportados
		CCPB= El Componente de Cobertura de la Población Potencialmente Beneficiaria.
2011 - 2014	$(CCC + CCPB^2) / 2$	CCC= El Componente de Calidad de diseño y Cumplimiento de metas CCPB= El Componente de Cobertura de la Población Potencialmente Beneficiaria.

Fuente: Elaboración propia con datos de GESOC (2008-2014).

Por último, a partir de 2011, GESOC rediseña el método de cálculo del INDEP e integra dos nuevos subíndices: el Componente de Calidad de diseño y Cumplimiento de metas (CCC); y el Componente de Cobertura de la Población Potencialmente Beneficiaria (CCPB).

Índice de Desempeño de los Programas Públicos (INDEP) 2008 - 2009

El Componente de Avance en los Indicadores reportados (CAI). Reporta el porcentaje de cumplimiento de la meta y qué tan bien fue definida.

Castigando tanto los porcentajes mediocres, iguales o menos de 50%, como los que exceden la meta.

- Definir metas retadoras y alcanzables (castigo si exceden el 125%, -10% cuando exceden hasta 150%, más -15% hasta 200%, más -20% cuando superan 200%)
- Cumplir con el 100% de las metas establecidas (los avances se dividen entre 10)

EL Componente de Fortaleza del sistema de Indicadores (CFI). Determinan la amplitud de la información brindada por los indicadores, a partir del número de dimensiones de desempeño reportadas.

- Utilizar indicadores que reflejen sus seis dimensiones de desempeño para demostrar el adecuado cumplimiento de sus metas: insumo, actividad, producto, resultado, impacto y eficiencia (costo/beneficio).
- Reportar información en todos sus indicadores.

$$\text{INDEP} = \text{CAI} \times \text{CFI}$$

Donde,

CAI= (Media ajustada de avance de indicadores) – (Castigo por diseño estratégico de metas)

CFI= (Factor de dispersión * 0.8) + (Factor de información sobre indicadores * 0.2) + (Premio a indicadores de relación * 0.16)

Factor de dispersión= $(1 - ((\% \text{Iinsumo})^2 + (\% \text{Iproceso})^2 + (\% \text{Iproducto})^2 + (\% \text{Iresultado})^2 + (\% \text{Iimpacto})^2))$.

Factor de información sobre indicadores= Porcentaje de indicadores incluidos en la EED con avance reportado.

Premio a indicadores de relación= $((1-(\%I_{\text{insumo}})^2 + (\%I_{\text{proceso}})^2 + (\%I_{\text{producto}})^2 + (\%I_{\text{resultado}})^2 + (\%I_{\text{impacto}})^2 + (\%I_{\text{insumo-producto}})^2 + (\%I_{\text{efectividad}})^2) - \text{el cálculo de dispersión con cinco categorías.}$

Semaforización INDEP,

Semáforo	Rojo	Amarillo	Verde
Selección adecuada de metas. Número de indicadores con avance mayor al 125%	Más de 2	1 o 2	ninguno
Fortaleza de los indicadores	Concentración de algún tipo de indicador	Distribución homogénea con concentración baja	Distribución homogénea
Reporte de la información de los indicadores	Menos del 60%	Entre 60 y 79%	Más del 80%

INDEP 2010

El Componente de Avance en los Indicadores reportados (CAI). Informa el grado de cumplimiento de todas las metas estratégicas y de gestión de cada programa, y cómo fueron reportadas. Y busca castigando tanto los porcentajes mediocres, iguales o menos de 50%, como los que exceden la meta.

- Promedio de avance de los indicadores de Fin y Propósito (Indicadores de resultados)
- Promedio de avance de los indicadores de Componente y Actividades (Indicadores de gestión)
- Castigo a indicadores de resultados con metas que exceden el 125% (-5% hasta 150%, o -10% cuando exceden el 150%)
- Castigo a indicadores de gestión con metas que exceden el 125% (-5% cuando exceden el 150%)

El Componente de Cobertura de la Población Potencialmente Beneficiaria (CCPB¹). Refleja el nivel de cobertura de la población potencialmente beneficiaria del programa, tal y como fue reportada en su Evaluación Específica de Desempeño (EED).

- Se calcula el avance de cobertura en función de la población potencial reportada entre la atendida
- Se realiza el mismo procedimiento sustituyendo la población potencial por la población objetivo
- Se calcula el promedio entre los cálculos obtenidos y se multiplica por 100 para elevar la calificación

$$\text{INDEP} = \text{CAI} + \text{CCPB}^1 / 2$$

Donde,

CAI= (Promedio de los Indicadores de Resultados) + (Promedio de los Indicadores de Gestión) – (Castigo por reporte de avances excesivos en de metas)

$$\text{CCPB}^1 = [(((\text{Población potencial} / \text{Población atendida}) + (\text{Población objetivo} / \text{Población atendida})) / 2) * 100]$$

Semaforización INDEP,

Semáforo	Rojo	Amarillo	Verde
Cumplimiento de aspectos susceptibles de mejora	Menos del 60%	Entre 60% y 79%	Más del 80%
Conclusiones del evaluador externo sobre el desempeño del programa	Menos del 60%	Entre 60% y 79%	Más del 80%

El Componente de Calidad de diseño y Cumplimiento de metas (CCC). Se integra por dos subcomponentes:

- Subcomponente de calidad de diseño (S-CaD). Informa el grado de calidad de diseño y la coherencia de cada programa, entendidos como el grado en que el programa identifica y sustenta con evidencia el problema que atiende y sus causas; establece una teoría de cambio y utiliza instrumentos de política consistentes; establece objetivos, metas e indicadores pertinentes; cuenta con las previsiones de implementación, seguimiento y evaluación correctas; así como de incorporación de las necesidades y expectativas de los ciudadanos que atiende. Se alimenta de las Evaluaciones de Consistencia y Resultados (ECRs) y de Diseño.

Con base en los términos de referencia establecidos por CONEVAL se buscó generar una metodología uniforme estandarizada. Para las Evaluaciones de Consistencia y resultados se analizó un *cuestionario de 51 preguntas que debía ser respondido por los evaluadores externos*. Del total de preguntas, 34 se encontraban bajo un formato binario (esto es, aceptaban una respuesta Sí/No), además de un esquema por niveles (a partir de 1 y hasta 4 dependiendo de si el programa cumplía con las características formuladas por CONEVAL para considerar una máximo cumplimiento o bien una disminución de esta calificación dependiendo de las características con las que contara el Programa). El resto de las preguntas tenían un esquema de respuesta abierta.

- Subcomponente de avance en los indicadores reportados (S-CAI). Informa el grado de cumplimiento de todas las metas estratégicas y de gestión de cada programa, tal como fueron reportadas. Se castiga a los programas que reportan avances muy superiores al 100% :
 - Indicadores nivel Fin, -5 entre 100% hasta 300%; o -10 más de 300%)
 - Indicadores nivel Propósito, -5 entre 115 hasta 299%; -10 si es igual o mayor 300%

- Indicadores de gestión, -5 entre 125 hasta 299%; -10 si es igual o superior 300%

El Componente de Cobertura de la Población potencialmente Beneficiaria (CCPB²). Este componente refleja el nivel de cobertura de la población potencialmente beneficiaria de cada programa.

$$\text{INDEP} = (\text{CCC} + \text{CCPB}^2) / 2$$

Donde,

$$\text{CCC} = (\text{S-CaD} + \text{S-CAI}) / 2$$

$$\text{CCPB}^2 = [(((\text{Población atendida} / \text{Población potencial}) + (\text{Población atendida} / \text{Población objetivo})) / 2) * 100]$$

$$\text{S-CaD} = \text{S-CAD-ECRs} + \text{S-CaD-EDs}$$

$$\text{S-CAI} = (\text{Promedio Avance de Indicadores de Fin} + \text{Promedio de Avance de Indicadores de Propósito} + \text{Promedio de Avance de Indicadores de Gestión}) - \text{Castigo por diseño } ad \text{ hoc.}$$

$$\text{S-CaD-ECRs} = (\text{Subíndice de Diseño y Alineación Estratégica} + \text{Subíndice de Operación} + \text{Subíndice de Orientación a resultados y a la ciudadanía})$$

$$\text{S-CaD-EDs} = (\text{Subíndice de Justificación del programa y contribución a los objetivos nacionales y a los sectoriales} + \text{Subíndice de Población potencial y objetivo} + \text{Subíndice de Matriz de Indicadores para Resultados}).$$